Title
A chimeric Ty3/Moloney murine leukemia virus integrase protein is active in vivo.

Journal
Journal of Virology, 72(5)

ISSN
0022-538X

Authors
Dildine, Sandra L
Respess, James
Jolly, Doug
et al.

Publication Date
1998-05-01

DOI
10.1128/jvi.72.5.4297-4307.1998

JOURNAL OF VIROLOGY, May 1998, p. 4297–4307 Vol. 72, No. 5
0022-538X/98/$04.00+0

A Chimeric Ty3/Moloney Murine Leukemia Virus Integrase Protein Is Active In Vivo

SANDRA L. DILDINE,[1] JAMES RESPESS,[2] DOUG JOLLY,[2] AND SUZANNE B. SANDMEYER[1]*

Department of Biological Chemistry, University of California—Irvine, Irvine, California 92697-1700,[1] *and Center for Gene Therapy, Chiron Technologies, San Diego, California 92121*

Received 14 November 1996/Accepted 26 January 1998

This report describes the results of experiments to determine whether chimeras between a retrovirus and portions of Ty3 are active in vivo. A chimera between Ty3 and a Neo^r-marked Moloney murine leukemia virus (M-MuLV) was constructed. The C-terminal domain of M-MuLV integrase (IN) was replaced with the C-terminal domain of Ty3 IN. The chimeric retroviruses were expressed from an amphotrophic envelope packaging cell line. The virus generated was used to infect the human fibrosarcoma cell line HT1080, and cells in which integration had occurred were selected by G418 resistance. Three independently integrated viruses were rescued. In each case, the C-terminal Ty3 IN sequences were maintained and short direct repeats of the genomic DNA flanked the integration site. Sequence analysis of the genomic DNA flanking the insertion did not identify a tRNA gene; therefore, these integration events did not have Ty3 position specificity. This study showed that IN sequences from the yeast retrovirus-like element Ty3 can substitute for M-MuLV IN sequences in the C-terminal domain and contribute to IN function in vivo. It is also one of the first in vivo demonstrations of activity of a retrovirus encoding an integrase chimera. Studies of chimeras between IN species with distinctive integration patterns should complement previous work by expanding our understanding of the roles of nonconserved domains.

Efficient retroviral vectors have played a central role in the development of gene therapy. One of the limitations of retroviral vectors, however, is the relatively random selection of insertion sites. This can result in disruption of the target genome and cause expression of the therapeutic gene to be unpredictable. The yeast retrovirus-like element Ty3 inserts with position specificity at the site of transcription initiation by RNA polymerase III (pol III). Potential limitations of retrovirus-based vectors could be resolved by Moloney murine leukemia virus (M-MuLV)-based retroviral vectors with the position-specific integration properties of the yeast retrovirus-like element Ty3. In this study, we constructed a retroviral vector with a chimeric integrase (IN) and determined its ability to function in vivo.

Retroviruses integrate throughout the genomes of their hosts. Although integration is not completely random, the mechanisms that determine the positions of integration are poorly understood (16, 47, 56). Factors influencing the structure of DNA appear to play a role in target site selection. Assembly of DNA into nucleosomes created favored sites for integration at positions where the major groove is on the exposed face of the nucleosomal DNA helix (54). A more detailed analysis of integration sites in DNA assembled into chromatin showed that DNA that is most severely distorted and that has a wider major groove within the nucleosome is a preferential target for integration (52). Bending of the target DNA by different DNA binding proteins or by phased tracts of adenosine residues can create favored integration sites in the region where the DNA is distorted (6, 48, 53). Another factor in target site selection is sequence- or structure-specific DNA binding proteins. Fusion proteins have been created between the IN protein of human immunodeficiency virus (HIV) and the DNA binding domains of λ (8), Lex A repressor (27), Zif268 (10), or avian sarcoma virus (ASV) IN and the DNA binding domains of the Lex A repressor (35). These fusion proteins were able to target integrations to regions surrounding the protein recognition sequences in in vitro assays. In naturally occurring interactions, the DNA binding protein(s) itself may promote integration by interacting with the integration machinery. For example, integration was stimulated by a specific interaction between HIV IN and a putative transcription activator (33). The most compelling example of such an interaction is the position-specific integration of the yeast retrovirus-like element Ty3. Integration of Ty3 adjacent to tRNA genes requires RNA pol III transcription factors, suggesting that position-specific integration may be influenced by an interaction between the Ty3 integration machinery and pol III transcription factors (38).

The yeast retrovirus-like element Ty3 is more closely related to the *Drosophila melanogaster* gypsy-like retroviruses and to animal retroviruses than to the other yeast retrotransposons (29). Ty3 is composed of a 4.7-kb internal domain flanked by 340-bp long terminal repeats (LTRs) (14). It is distinguished from other retrotransposons and from retroviruses by its unique integration specificity. De novo insertions of Ty3 elements into yeast genomic DNA were shown to be integrated within 1 to 2 bp of the site of initiation of transcription of tRNA genes (12). Subsequently it was shown that the RNA pol III-transcribed genes, 5S and U6, can also serve as specific targets for Ty3 integration (13). The tRNA gene target must be transcriptionally competent since promoter mutations that abolish transcription prevent integration (13). Experiments using an in vitro integration assay and fractionated transcription extracts showed that transcription factors TFIIIB and TFIIIC are required for integration but that RNA pol III is not (15, 38). Ty3 integration does not significantly affect the expression of the adjacent tRNA gene (37), suggesting that Ty3 may have evolved naturally to insert in a nondetrimental position in the yeast genome. If the integration specificity of Ty3 could be

* Corresponding author. Mailing address: Department of Biological Chemistry, College of Medicine, University of California—Irvine, Irvine, CA 92697-1700. Phone: (714) 824-7571. Fax: (714) 824-2688. E-mail: sbsandme@uci.edu.

adapted to retrovirus-based gene therapy vectors, this would result in a therapeutic vector that integrates into a predictable site in the genome. In addition, tRNA genes are redundant in the human genome and, because they are expressed constitutively, are likely to be located in accessible regions of chromatin in many cell types. Therefore, integration adjacent to a tRNA gene may lead to more predictable levels of expression of the therapeutic vectors than integration into random sites.

Studies on retroviral IN proteins suggest there are domains that can be separated to generate functional IN chimeras. Computer alignment of the amino acid sequence of retroviral and retrotransposon IN proteins show that there is a highly conserved region which includes seven invariant residues [an HHCC metal finger and a DD(35)E active-site motif], flanked by N-terminal and C-terminal domains (31, 36). The central core region beginning C terminal to the HHCC motif appears to constitute a domain by structural and functional criteria. Limited proteolysis of HIV IN showed that a core of about 120 amino acids (aa) including the DD(35)E motif was relatively resistant to proteolysis (22). Expression of recombinant subclones of HIV type 1 further showed that the region from aa 50 to 186 containing the same subset of conserved residues was sufficient to carry out the disintegration reaction (9, 65), indicating that this domain functions independently in polynucleotidyl transfer. The minimal domain for disintegration activity of M-MuLV IN includes the DD(35)E motif and most of the C-terminal region (32). The region containing the HHCC and DD(35)E motifs, conserved among all retrovirus IN proteins, shares approximately 25% amino acid identity between M-MuLV and Ty3 (data not shown). Amino acid substitutions in the DD(35)E catalytic triad block retroviral integration in vivo (11, 42, 60) and in vitro (21, 22, 40, 43, 64). Ty3 IN also requires the conserved DD(35)E motif, since amino acid substitutions in the active site blocked 3′-end processing in vivo (39) and virus-like particles containing the IN mutations were unable to catalyze integration in vitro (38). The N-terminal (to the HHCC) and C-terminal domains are poorly conserved among retroviruses. The C-terminal domain shows the greatest variability in size and sequence (31, 36). The C-terminal domains of HIV, M-MuLV, and Ty3 IN are about 100, 140, and 230 aa, respectively. The C terminus of retroviral IN proteins contains a domain that has been shown to have nonspecific DNA binding activity (24, 36, 49, 58, 65, 67), but the DNA binding domain is not required for catalytic activity. Its significantly larger size in Ty3 suggests that it could perform other functions. This domain in Ty3 is a candidate for targeting integration. We have replaced the C domain of M-MuLV IN with the C domain of Ty3 IN to determine whether portions of Ty3 IN could substitute for portions of M-MuLV IN and, if so, whether changes in patterns of M-MuLV integration result.

Three independently integrated chimeric retroviruses containing the Ty3 C-terminal domain ($A_MB_MC_T$; see the legend to Fig. 1 for nomenclature) were identified. Sequence analysis of each of these chimeric viruses has revealed the maintenance of the Ty3 C-terminal sequences. Short direct repeats of the flanking genomic DNA were detected, indicating that these are true integrations. Searching the National Center for Biotechnology Information (NCBI) sequence database with the rescued flanking sequence did not reveal the presence of a tRNA gene. Therefore, these integrations do not appear to be position specific. These results show that the C-terminal domain of Ty3 IN can substitute for M-MuLV sequence and provide some IN activity.

MATERIALS AND METHODS

DNA constructs. Wild-type M-MuLV retroviral vector plasmids used in this study are designated pRgpNeo and pRgpKan. pRgpNeo is a rescuable shuttle vector, derived from the BAG vector (51) in which the β-galactosidase sequences, bp 812 (*Aat*II) to 4156 (*Sal*I), were replaced with M-MuLV *gag-pol* sequences, bp 367 (*Aat*II) to 5872 (*Sca*I) (61), from pMLV-K (46). The vector expresses G418 resistance, and *gag* and *pol* gene products encapsidate the vector RNA in mammalian cells. In a bacterial host, pRgpNeo exists as a plasmid containing the ColE1 origin of replication and expresses kanamycin resistance. pRgpKan is similar to pRgpNeo but contains a single LTR and no polyomavirus sequences. When pRgpKan is transiently expressed in mammalian cells along with an envelope gene, a retroviral vector identical to pRgpNeo is produced.

The chimeric retroviral vector $pRgpA_MB_MC_T$ was constructed via a multistep cloning process. M-MuLV nucleotides 3706 to 6538 from p2XMLV (50) encompassing IN were cloned into the pIBI-20 (17) phagemid vector cleaved in the polylinker with *Sal*I and *Bam*HI to create pMLVIN. Ty3 IN sequences were provided by pVB193 (4), which contains Ty3 sequences from 3132 to 5351 (30) encompassing IN in the pIBI-20 phagemid vector. A 4.4-kb *Sca*I fragment containing IN sequences from pMLVIN was ligated to a 4.5-kb *Sal*I-*Sca*I fragment containing Ty3-1 IN sequences from pVB193 to generate plasmid pMLV/Ty3IN. The phagemid vector pMLV/Ty3IN contains the M-MuLV IN- and Ty3 IN-coding sequences in tandem. It served as a template for the synthesis of single-stranded DNA for mutagenesis. The 41-mer oligonucleotide, 5′-CCTAAATCA ATTTCAAATGGGGTGAGGCCATGGGGGCCCGG-3′, complementary to M-MuLV nucleotides 5379 to 5399 and Ty3 nucleotides 4362 to 4381 was used to loop out the intervening sequences and join the M-MuLV B domain to the Ty3 C domain via single-stranded oligonucleotide mutagenesis (41). This generated plasmid $pA_MB_MC_T$. The 44-mer oligonucleotide, 5′-GGATGTTTCGG GGGTTATAGTTAAAAGATACTCCTCCCATCTCC-3′, complementary to M-MuLV nucleotides 4597 to 4619 and Ty3 nucleotides 3450 to 3470 was used to loop out the intervening sequences and join the coding region for the first 4 aa of M-MuLV IN to the coding region for the N terminus of Ty3 IN. This generated plasmid $pA_TB_TC_T$. The 44-mer oligonucleotide, 5′-CTTCTAGGGA ATAATTCTTTCGTTTTCCTTGGTAGACCCAATAC-3′, complementary to M-MuLV nucleotides 4712 to 4733 and Ty3 nucleotides 3633 to 3654 was used to loop out the intervening sequences and join the M-MuLV A domain to the Ty3 B domain. This generated plasmid $pA_MB_TC_T$.

The chimeric IN sequences were cloned into the retroviral vector by ligation of a 4.1-kb *Nhe*I-*Sal*I fragment from pRgpKan, which encompassed the M-MuLV LTR, *gag*, and part of *pol*; a 3-kb *Nhe*I-*Sal*I fragment from BAGΔX (truncated version of the BAG vector in which the polyomavirus sequences have been deleted and there is only one LTR [51]), which provided the neomycin phosphotransferase resistance (Neo^r) gene from transposon Tn*5* under control of the simian virus 40 (SV40) promoter and ColE1 origin of replication; and an approximately 2.7-kb *Sal*I-*Xho*I fragment from $pA_MB_MC_T$, $pA_TB_TC_T$, or $pA_MB_TC_T$, which provided the chimeric IN sequences. In addition to Ty3 IN sequences, the 2.7-kb *Sal*I-*Xho*I fragment contains the majority of the Ty3 LTR sequences to bp 5332 (30). Ligation products were transformed into *Escherichia coli* HB101 via electroporation. Transformants were selected by growth in the presence of kanamycin (50 μg/ml). This generated the chimeric retroviral vector plasmids $pRgpA_MB_MC_T$ (two independent clones, 3-7 and 4-11) and $pRgpA_TB_TC_T$ and $pRgpA_MB_TC_T(-)$, which contained the chimeric IN sequences inserted in the opposite orientation. The presence of the chimeric IN sequences was verified by restriction enzyme digestion and sequence analysis in the region of the M-MuLV-Ty3 junctions.

Cell lines. The cell line 293 2-3 (7) is derived from the human adenovirus type 5-transformed embryonal kidney cell line 293 (ATCC CRL1573) and expresses M-MuLV *gag* and *pol* genes. The packaging cell line NC10 is derived from the human fibrosarcoma cell line HT1080 (ATCC CCL121). NC10 cells express the M-MuLV 4070A amphotrophic envelope (5). All cells were grown in Dulbecco's modification of Eagle's medium supplemented with 10% fetal calf serum (Irvine Scientific).

Virus production and G418 selection for integration. The retroviral vectors were independently cotransfected (28) into 293 2-3 cells along with pMLP-G at a 1:1 ratio (10 μg of each). Plasmid pMLP-G expresses vesicular stomatitis virus (VSV) G protein and was used to pseudotype retroviral vector particles (7). After 48 h, 10 ml of medium from transfected cell cultures was collected, filtered through 45-μm-pore-size cellulose acetate filters (Nalgene Inc.), and placed onto NC10 cells. After 24 h, the transduced NC10 culture was subjected to G418 (900 μg/ml) selection for 24 to 48 h. Selection was continued at a reduced level of G418 (600 μg/ml) until nontransduced control cultures no longer contained viable cells. The resulting NC10 culture produced retroviral vectors consisting of (i) the chimeric RNA genome and Gag and Gag-Pol from the chimeric construct and (ii) amphotropic envelope. NC10 producer cells were grown to confluency. Medium was changed and collected after 24 h. Ten milliliters of filtered medium was placed onto target HT1080 cells at approximately 50% confluency. After 24 h, medium was replaced with fresh Dulbecco modified Eagle medium plus 10% fetal calf serum. After an additional 24 h, the cells were placed under G418 selection at 900 μg/ml for 24 to 48 h. Selection continued at 600 μg/ml until nontransduced control cultures no longer contained viable cells. G418-resistant cell cultures were expanded and used to prepare genomic DNA.

Genomic DNA extraction, Southern analysis, and rescue of integrated retroviral vectors. High-molecular-weight DNA was prepared from G418-resistant HT1080 cells infected with the chimeric $RgpA_MB_MC_T$ and wild-type RgpNeo retroviral vectors. Approximately 10 μg of each sample of genomic DNA was digested with the restriction enzyme *Pst*I, and the DNA was electrophoresed on a 0.9% agarose gel. The DNA was transferred onto Zeta-Probe GT nylon membranes (Bio-Rad) by using a PosiBlot pressure blotter (Stratagene Inc.). The blots were hybridized at 42°C in the presence of 50% formamide with a fragment of M-MuLV (nucleotides 4643 to 5873) or Ty3 (nucleotides 3132 to 5332) which was specific for the respective IN-coding regions. The probe was synthesized by extending random primers in the presence of [α-^{32}P]dATP with the Megaprime DNA labeling system (Amersham, Inc.). Integrated retroviral vectors were recovered by digestion of 10 μg of genomic DNA with *Sca*I, extraction of DNA with phenol-chloroform, and precipitation of DNA with ethanol. *Sca*I-digested genomic DNA was ligated with T4 DNA ligase (New England BioLabs, Inc.). DNA was extracted with phenol-chloroform and precipitated with ethanol precipitation in the presence of 10 μg of glycogen. DNA was resuspended in 7 μl of TE, (10 mM Tris base, 1 mM EDTA [pH 8.0]). One microliter was used to quantitate the DNA by fluorometry using a Mini TKO 100 DNA Fluorometer (Hoefer Scientific). The remainder of the DNA (6 μl) was used to transform *E. coli* DH12S (Life Technologies) via electroporation with a Gene Pulser (Bio-Rad). *E. coli* transformants were grown at 30°C and plated onto LB medium containing 50 μg of kanamycin per ml. Rescued retroviral vector plasmids were recovered by the alkaline lysis procedure (3). The retroviral vector and flanking genomic sequences were subcloned by standard DNA cloning techniques (3). In general, the rescued plasmids were digested with *Nhe*I and fragments were separated by electrophoresis in agarose gels. The *Nhe*I fragments representing the retroviral vectors were circularized. The *Nhe*I fragments representing the flanking genomic DNA were cloned into the *Xba*I site of the pIBI-20 vector.

Sequence analysis. Dideoxynucleotide sequencing was performed by the method of Sanger et al. (57), using the Sequenase enzyme (U.S. Biochemicals). Oligonucleotide primers used for sequence analysis were as follows: 394, 5′-ATGCATCTCTATGCAC-3′ (complementary to Ty3 nucleotides 5307 to 5323); 175, 5′-CAGGGTGACGTATTGTC-3′ (complementary to Ty3 nucleotides 5042 to 5054); 194, 5′-ATGCATCTCTATGCAC-3′ (complementary to Ty3 nucleotides 4581 to 4596); 199, 5′-CAACTGGCTCTAGAC-3′ (M-MuLV nucleotides 5317 to 5331); 387, 5′-GTCTCGCTGTTCCTTGGGAG-3′ (M-MuLV nucleotides 80 to 99); universal primer, 5′-GTAAAACGACGGCCAGTG-3′ (complementary to pIBI-20 nucleotides 341 to 358); and reverse primer, 5′-CAGGAAACAGCTATGACC-3′ (pIBI-20 nucleotides 202 to 219). Sequencing reactions were fractionated by electrophoresis in 8% polyacrylamide/bisacrylamide (19:1, National Diagnostics)–8 M urea gels and visualized by autoradiography.

PCR amplification of HT1080 preintegration genomic DNA. High-molecular-weight genomic DNA isolated from HT1080 cells was used as a substrate for asymmetric PCR using oligonucleotide primers 406 (5′-GAACCACTAAGTTTGCTTGTGG-3′) and 407 (5′-CCATGAGAAATACTAGGTGACTGC-3′). The sequences of oligonucleotide primers 406 and 407 represent opposite strands on the 5′ and 3′ flanks, respectively, of the rescued integration from the 4-11 retroviral chimeric clone 10 [4-11(10)]). One microgram of HT1080 genomic DNA was amplified with 50 and 0.5 pmol of oligonucleotide primers 406 and 407, respectively, 2 mM $MgCl_2$, and 2.5 U of *Taq* polymerase (Perkin-Elmer, Inc.). PCR products were collected by ethanol precipitation in the presence of 2.5 M sodium acetate and annealed to oligonucleotide 407 prior to sequence analysis.

Northern analysis. Total cytoplasmic RNA was extracted from NC10 producer cells as described previously (19). Approximately 10 μg of RNA from each sample was denatured by reaction with glyoxal as described previously (45) and subjected to electrophoresis in a 1.1% agarose gel in 10 mM NaP (pH 7) at ~130 V for 4 h. The RNA was transferred in a PosiBlot pressure blotter (Stratagene) to a Duralon-UV membrane and cross-linked in a UV Stratalinker 1800 (Stratagene). Samples on identical membranes were hybridized with a fragment of M-MuLV (nucleotides 4643 to 5750) or Ty3 (nucleotides 3132 to 5332) which was specific for the respective IN-coding regions. Membranes were stripped of probes by the addition of boiling 0.1× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–1% sodium dodecyl sulfate and rehybridized with a fragment from M-MuLV *gag-pol* sequences (nucleotides 1906 to 4444). Membranes were stripped again and rehybridized with a fragment from the Neo^r gene (*Hin*dIII to *Eco*RI). The probes were synthesized by extending random primers in the presence of [α-^{32}P]dATP with the Megaprime DNA labeling system (Amersham). Membranes were hybridized to probes and washed as described previously (14) except that the hybridization and second wash were done at 42°C.

RESULTS

Activity of retroviral vectors containing chimeric IN. The C-terminal region of retroviral IN has been implicated in binding of target DNA. To test whether this region of the position-specific Ty3 IN could function in a retroviral context, we constructed chimeras based on a M-MuLV retroviral vector, pRgpKan (Fig. 1). pRgpKan contains a single LTR and *gag* and *pol* genes from M-MuLV, the Neo^r gene from transposon Tn*5* driven by an SV40 promoter, and a bacterial origin of replication. The single LTR directs both initiation and termination of transcription so that an RNA genome is transcribed with LTR information at both ends. The wild-type M-MuLV retroviral vector used in these experiments was pRgpNeo, derived from the BAG vector (51). Plasmid pRgpNeo differs from pRgpKan by having two LTRs and sequences from the polyomavirus early region. Once pRgpKan is expressed in mammalian cells, an RNA genome which is identical to pRgpNeo is transcribed; therefore, viruses produced from these two retroviral vectors are identical. The retroviral vector $pRgpA_MB_MC_T$, substituting the C domain of the Ty3 IN for the C domain of M-MuLV IN in the pRgpKan vector, was constructed as described above. To assay the function of the chimeric IN protein, retrovirus containing chimeric IN was generated via a three-step process in cell culture. The chimeric retroviral vectors were cotransfected into 293 2-3 cells (7) along with pMLP-G (expresses the VSV G protein) at a 1:1 ratio. In 293 2-3 cells, the retroviral vector RNA genome was encapsidated along with Gag and Gag-Pol into viral particles. Budded particles contained VSV G protein on the outer surface. Filtered supernatant fluid containing these particles was transferred onto NC10 cells. The transduced NC10 culture was placed under G418 selection. NC10 cells were maintained under G418 selection until nontransduced control cultures no longer contained viable cells. As judged by the relative number of G418-resistant cells, the chimeric genomes were packaged at the same rate as observed for the wild-type pRgpNeo vector (data not shown). The resulting culture then produced retrovirus consisting of the chimeric RNA genome, Gag and Gag-Pol expressed from the chimeric construct, and M-MuLV amphotrophic envelope.

Activity of the chimeric IN protein was assayed by placing filtered supernatants from the NC10 producer cells onto the target, HT1080 cells (~10^6 cells) and selecting for G418-resistant transductants. Table 1 lists the approximate number of G418-resistant cell colonies per 10 ml of filtered producer supernatant fluid. Producer supernatants from the two independent clones, 3-7 and 4-11, of the $pRgpA_MB_MC_T$ chimeric construct were each assayed twice on target HT1080 cells. The $RgpA_MB_MC_T$ chimeric retrovirus yielded one to three G418-resistant cell colonies. In contrast, infection with the wild-type RgpNeo retrovirus yielded more than 10^4 G418-resistant cell colonies. The retroviral vector $pRgpA_MB_TC_T(-)$, containing chimeric IN sequences cloned into the retroviral vector in the reverse orientation, was used as a control. As expected, this construct, which did not have a functional IN protein, did not yield any G418-resistant cells. Supernatant from the NC10 producer cell line alone also did not yield any G418-resistant cells.

Maintenance of Ty3 IN sequences in integrated retroviral vector DNA. To determine whether the chimeric IN sequences were maintained in the G418-resistant HT1080 cells, genomic DNA was isolated from HT1080 cells, digested with *Pst*I, and subjected to Southern analysis as shown in Fig. 2. *Pst*1 cleaves within the retroviral vector on either side of the IN sequences to liberate a 6-kb (pRgpKan) or 6.5-kb ($pRgpA_MB_MC_T$) fragment (Fig. 2C). Identical Southern blots were hybridized with a M-MuLV or Ty3 IN-coding sequence specific α-^{32}P-labeled probes (Fig. 2A and B). Genomic DNA from RgpNeo-infected, G418-resistant HT1080 cells hybridized with the M-MuLV probe (Fig. 2A, lane 3) but not the Ty3 probe (Fig. 2B, lane 3). Genomic DNAs from $RgpA_MB_MC_T$-infected, G418-

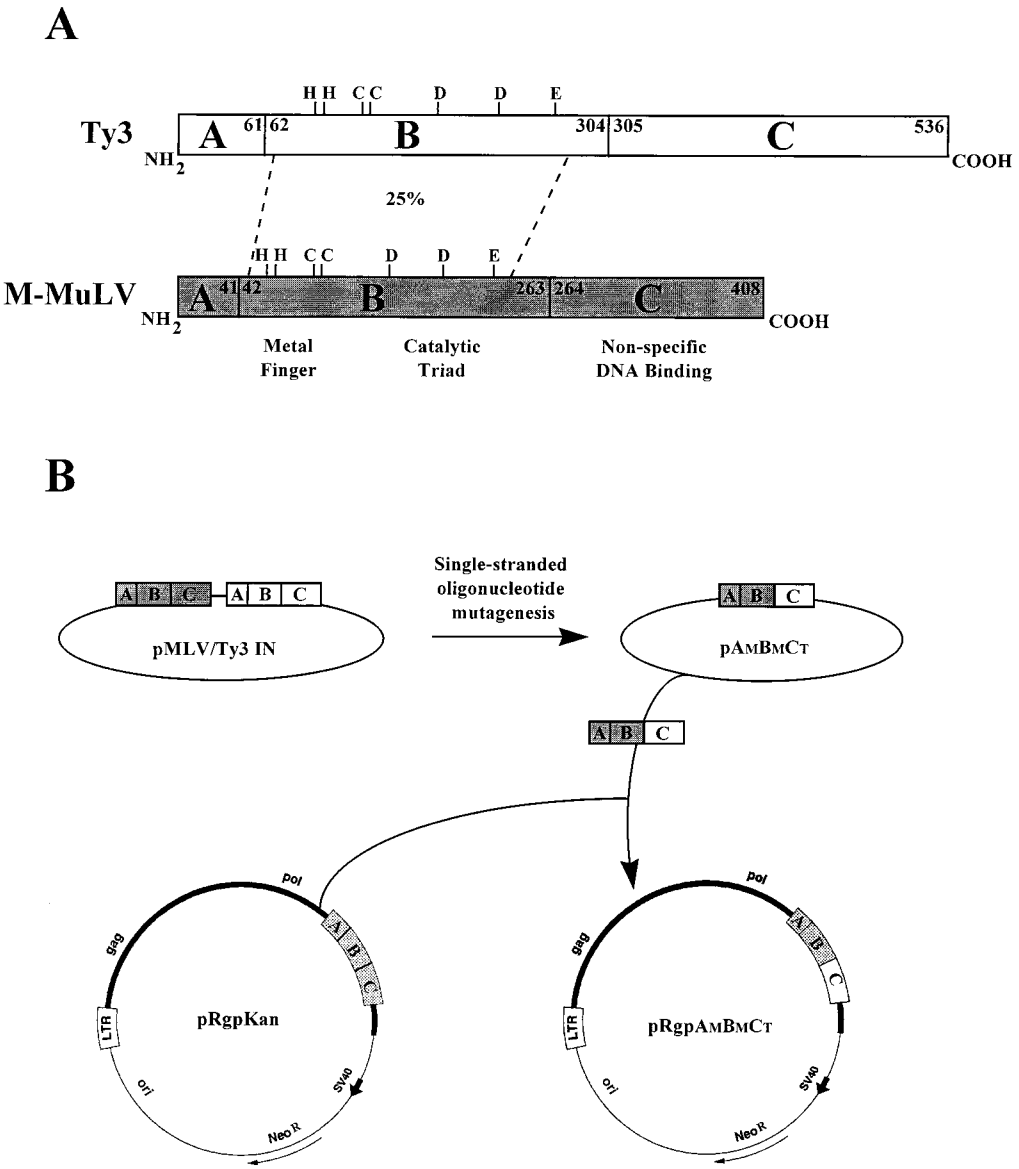

FIG. 1. (A) Primary structure comparison of Ty3 and M-MuLV IN proteins. Open and filled boxes represent Ty3 and M-MuLV IN amino acids, respectively. IN proteins are divided into A (amino-terminal), B (central or core), and C (carboxyl-terminal) domains. Domains were chosen based on amino acid similarity and functional and structural similarities. The numbers in the boxes indicate the amino acid boundaries of the domains. H and C indicate conserved histidine and cysteine, respectively, amino acids which constitute a metal finger domain. D, D, and E indicate conserved aspartic acid and glutamic acid, respectively, amino acids which constitute the catalytic triad of the protein(s). (B) Construction of chimeric retroviral vectors. Single-stranded DNA from pMLV/Ty3IN was synthesized and served as the template for single-stranded oligonucleotide mutagenesis to produce $pA_MB_MC_T$, where A, B, and C refer to the IN domains listed above, $_M$ refers to M-MuLV-derived sequences, and $_T$ refers to Ty3-derived sequences. Plasmid pRgpKan is a rescuable shuttle vector containing the M-MuLV LTR, *gag* and *pol* genes, and SV40 promoter (SV40) driving a Neo^r gene and a ColE1 origin of replication (ori). Plasmid $pRgpA_MB_MC_T$ contains the chimeric IN sequences in place of the wild-type M-MuLV IN sequences in a pRgpKan background.

resistant HT1080 cells from clone 3-7 (Fig. 2A and B, lanes 4) and clone 4-11 (Fig. 2A and B, lanes 5) hybridized with both the M-MuLV- and Ty3-specific probes, as expected. However, the size of the hybridizing fragment from genomic DNA of clone 3-7-infected cells was larger than expected. *Pst*I-digested genomic DNA from control HT1080 cells (lanes 2) did not hybridize with either probe. The expected size of the fragments was determined by *Pst*I digestion of plasmids pRgpKan (lanes 6), $pRgpA_TB_TC_T$ (lanes 7), and $pRgpA_MB_MC_T$ (lanes 8).

Rescue of integrated chimeric retroviral vectors. To verify the maintenance of the Ty3 IN-coding sequences and determine the sequence of the genomic DNA at the insertion site, integrated retroviral vector DNA was isolated from the genomic DNA. Figure 3 outlines the procedure used to rescue the integrated retroviral vector DNA and flanking genomic DNA. This procedure had the advantage of recovering both the 5′ and 3′ flanking genomic DNA; thus, whether a target site repeat exists, which is indicative of an integration event, can be determined. Genomic DNA was isolated from G418-resistant cells containing integrated chimeric retroviral vectors. The genomic DNA was digested with the restriction enzyme *Sca*I, which does not cut within the retroviral vector sequences. It was ligated at low plasmid concentrations, which favor self-ligation, and transformed into *E. coli*. Transformants were

TABLE 1. Transduction of HT1080 target cells with chimeric IN-containing retroviral vectors

Construct	Isolate	Titer on HT1080 cells[a]
$pRgpA_MB_MC_T$	3-7	0, ≥1
	4-11	2, ≥3
pRgpNeo		$>10^4$, $>10^4$
$pRgpA_MB_TC_T(-)$		0, 0
No vector		0, 0

[a] G418-resistant cell colonies surviving per 10 ml of filtered, NC10 producer cell supernatant. Numbers shown represent two separate transient transfections of 293 2-3 cells with the retroviral vectors and the establishment of two independent NC10 producer cell lines followed by titration on HT1080 cells.

selected by kanamycin resistance. To verify the rescue of a full-length retroviral vector, the rescued plasmids were digested with *Nhe*I, which cuts once in each LTR. Rescued plasmids containing full-length retroviral vectors should yield a 9.2-kb (pRgpNeo) or a 9.7-kb ($pRgpA_MB_MC_T$) fragment representing the retroviral vector sequence. Other fragments(s) should contain junction and flanking genomic DNA sequences. Table 2 lists the rescued plasmids which contain either the chimeric $pRgpA_MB_MC_T$ or wild-type pRgpNeo retroviral vector. Three independent integrations were rescued from cells infected with the 4-11 chimeric retroviral vector: 4-11(3), 4-11(9), and 4-11(10). The recovered plasmids contained the expected 9.7-kb retroviral vector *Nhe*I fragment and additional fragment(s) representing flanking genomic DNA. All rescued plasmids from the 3-7 chimeric clone were similar to 3-7(19), which contained the 9.7-kb *Nhe*I fragment but no fragments representing flanking genomic DNA. If the rescued plasmid is too large for maintenance in *E. coli*, the genomic sequences may be lost by recombination between the LTRs. In that event, only the retroviral vector sequence would be rescued. Three independent integrations were rescued from the cells infected with the wild-type RgpNeo retrovirus. Each of these contained the expected 9.2-kb retroviral vector *Nhe*I fragment and additional fragment(s) representing flanking genomic DNA. To compare the rescued retroviral vector sequences to the input retroviral vector sequences and to facilitate sequence analysis of the flanking genomic DNA, the rescued retroviral vector sequences and flanking genomic DNA were subcloned.

To subclone the rescued retroviral vector sequences and flanking DNA, the recovered plasmids were digested with *Nhe*I. The 9.2- or 9.7-kb fragment was isolated, circularized by ligation and transformed into *E. coli*. The reconstructed retroviral vectors are listed in Table 2. *Bgl*II cuts within the M-MuLV IN- and Ty3 IN-coding sequences to yield a distinct pattern when hybridized with either a M-MuLV IN- or Ty3 IN-coding sequence-specific probe (Fig. 4C). Rescued and input retroviral vector plasmids were digested with *Bgl*II and subjected to Southern analysis. Figure 4 shows the *Bgl*II digestion pattern of input and rescued retroviral vectors pRgpNeo, 3-7, and 4-11 when hybridized with either an M-MuLV IN (Fig. 4A)- or Ty3 IN (Fig. 4B)-coding sequence-specific probe. The rescued wild-type pRgpNeo vector yielded a hybridization pattern identical to that of the input plasmid (Fig. 4A and B; compare lanes 9 to 11 with lane 1). Each rescued plasmid from the 3-7 and 4-11 chimeric clones yielded a hybridization pattern identical to that of the input chimeric plasmids (Fig. 4A and B; compare lanes 5 to 8 with lanes 2 and 3). These results indicated that no gross deletion or rearrangements of the chimeric IN coding sequences had occurred. Sequence analysis of the entire Ty3 IN C domain in the three 4-11 chimeric clones verified the presence of the Ty3 IN-coding sequences, with the exception of single nucleotide changes identified in the 4-11(3) and 4-11(10) rescued vectors. The single nucleotide changes would result in an aa substitution of aspartic acid for asparagine at position 382 in 4-11(3) and lysine for glutamic acid at position 486 in 4-11(10) in the Ty3 IN C domain. The significance of these second-site mutations remains to be determined. Nonetheless, the maintenance of Ty3 IN sequences in the rescued retroviral vectors indicates that these integrations were mediated by a chimeric IN protein containing the C-terminal domain of Ty3 IN.

Sequence analysis of the flanking genomic DNA. A hallmark of retroviral integration into genomic DNA is the direct duplication of the target genomic DNA. The size of the target site repeat is determined by the distance between the positions of IN nicking on the two strands. The separation of nicks in the two strands varies among retroviruses; for M-MuLV it is 4 bp (62), and for Ty3 it is 5 bp (14). Flanking genomic DNA rescued from the three wild-type RgpNeo and three $RgpA_MB_MC_T$ chimeric retroviral vector integrations was subcloned into the pIBI-20 vector (Table 2) for sequence analysis. Flanking sequence from the pRgpNeo retroviral vector, as

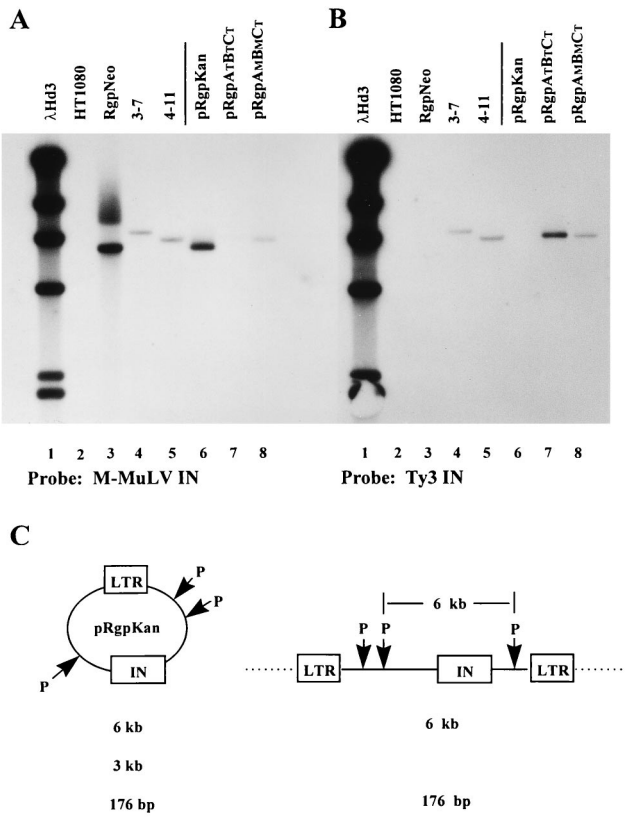

FIG. 2. Maintenance of chimeric IN sequences in genomic DNA from chimeric retroviral vector-infected HT1080 cells. (A and B) Southern analysis of genomic DNA isolated from G418-resistant HT1080 cells infected with pRgpNeo (lanes 3), $pRgpA_MB_MC_T$ clone 3-7 (lanes 4), or $pRgpA_MB_MC_T$ clone 4-11 (lanes 5). HT1080 genomic DNA is represented in lanes 2. Lanes 2 to 5 contain 10 μg of high-molecular-weight genomic DNA digested with restriction enzyme *Pst*I. *Pst*I-digested plasmids pRgpKan (lanes 6), $pRgpA_TB_TC_T$ (lanes 7), and $pRgpA_MB_MC_T$ (lanes 8) are included as molecular weight controls. Duplicate samples were analyzed via Southern hybridization with an α-^{32}P-labeled M-MuLV IN-specific probe (A) and an α-^{32}P-labeled Ty3 IN-specific probe (B). Lanes 1 contain λ *Hin*dIII as a size marker. (C) Schematic diagram of *Pst*I restriction enzyme sites on circular plasmid pRgpKan (left) and integrated pRgpNeo (right). The IN-specific probes should detect a 6-kb (pRgpNeo) or a 6.5-kb ($pRgpA_MB_MC_T$) *Pst*I fragment encompassing the IN sequences.

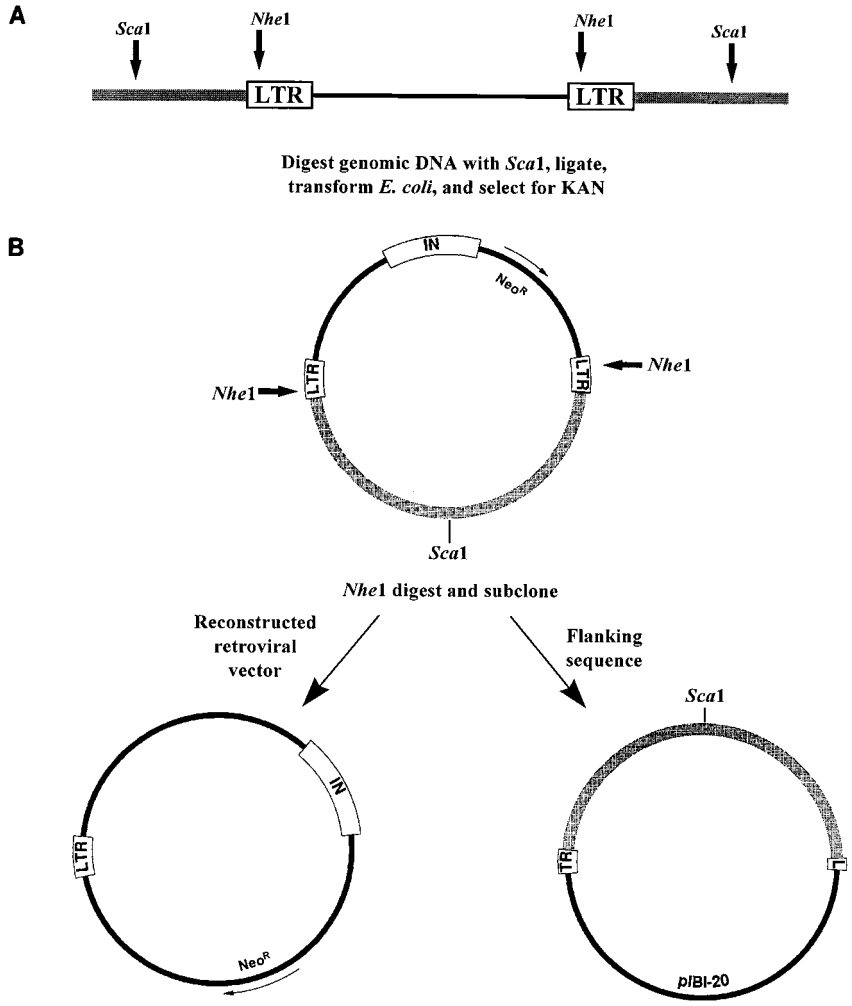

FIG. 3. Strategy to rescue integrated retroviral vector DNA along with target genomic DNA. High-molecular-weight genomic DNA was isolated from G418-resistant HT1080 cells infected with the retroviral vectors. (A) Schematic representation of retroviral vector DNA integrated into the genomic DNA. The open boxes represent the retroviral LTR sequences, the solid black line represents the internal retroviral vector sequence, and the stippled boxes represent the genomic DNA. Genomic DNA was digested with *Sca*I, circularized, and transformed into *E. coli* DH12S. Transformants were selected by kanamycin resistance. (B) The rescued plasmid. To facilitate further analysis, the genomic flanking sequence was subcloned into the pIBI-20 vector, and the retroviral vector sequence was reconstructed by *Nhe*I digestion of the rescued plasmid and circularization.

expected, revealed the presence of 4-bp target site repeats for all three rescued integrations (Fig. 5). Flanking sequence from each of the three integrations from the $RgpA_MB_MC_T$ chimeric retroviral vector contained target site repeats, indicating that the incorporation of chimeric retroviral sequences into genomic DNA occurred by integration. Interestingly, two of the integrations produced a 4-bp target site repeat, similar to that made by M-MuLV IN, while the other integration produced a 5-bp target site repeat, similar to that made by Ty3 IN (Fig. 5). To verify the 5-bp repeat, the preintegration genomic DNA was amplified and sequenced by asymmetric PCR (Fig. 6). The nucleotide sequence on the left represents the 3′ LTR, and the flanking DNA with the 5-bp repeat, AGGGT, is indicated. The nucleotide sequence on the right represents the preintegration genomic DNA containing the target AGGGT and flanking sequence. Below the sequencing gel, the nucleotide sequence of the preintegration genomic DNA is shown with the positions of the staggered cuts inferred to be made by the chimeric IN protein indicated by the arrows. Joining of 3′ ends of viral DNA at 5′-overhanging positions in target DNA followed by repair, presumably by host enzymes, results in a 5-bp duplication of the target site (underlined). The substitution of Ty3 sequences in the C-terminal domain of M-MuLV IN may result in an IN protein with hybrid strand transfer activity, or it may simply make strand transfer less precise.

TABLE 2. Rescued chimeric retroviral vectors and subclones

Input chimera	Rescued chimera	Size (kb)	
		Reconstructed retroviral vector	Subcloned flanking sequence
4-11	4-11(9)	9.7	11
	4-11(10)	9.7	12
	4-11(3)	9.7	1.0
			5.0
3-7	3-7(19)	9.7	—[a]
pRgpNeo	2	9.2	5.0
	3	9.2	5.5
			0.45
	10	9.2	1.8

[a] Clone contained no flanking sequence.

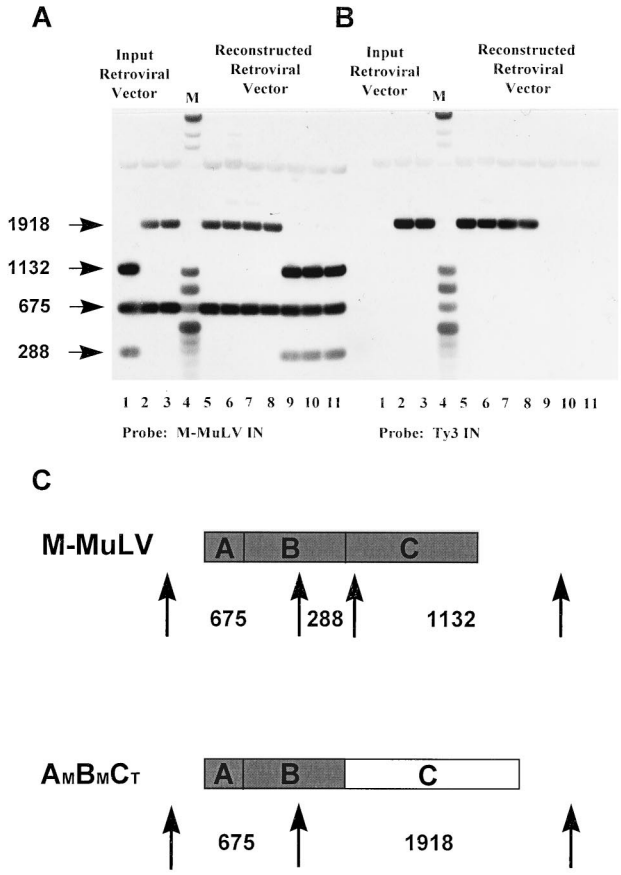

FIG. 4. Maintenance of chimeric IN sequences in rescued retroviral vectors. (A and B) Southern hybridization of subcloned, rescued retroviral vectors with an α-^{32}P-labeled M-MuLV (A) or Ty3 (B) IN-coding sequence specific probe. The subcloned retroviral vectors rescued from HT1080 cells infected with pRgpNeo (lanes 9 to 11), pRgpA$_M$B$_M$C$_T$ clone 3-7 (lanes 5), and clone 4-11 (lanes 6 to 8) were digested with restriction enzyme *Bgl*II, and restriction patterns were compared to the *Bgl*II digestion patterns of the input retroviral vector plasmids pRgpKan (lanes 1), pRgpA$_M$B$_M$C$_T$ clone 3-7 (lanes 2), and clone 4-11 (lanes 3). Lambda DNA digested with *Hin*dIII and DNA size markers (no. VIII; Boehringer Mannheim Inc.) were used as molecular weight markers (M; lanes 4). The size, in base pairs, of the expected *Bgl*II fragments are indicated at the left. (C) *Bgl*II restriction enzyme sites in pRgpNeo and M-MuLV and Ty3 IN sequences are indicated by the vertical arrows. *Bgl*II cuts within the pRgpKan and pRgpA$_M$B$_M$C$_T$ IN sequences to yield fragments of the indicated sizes.

Integration of the RgpA$_M$B$_M$C$_T$ chimeric retroviral vector is not adjacent to a tRNA gene. To determine whether the chimeric IN protein with the altered strand transfer activity was targeting integration to a tRNA gene, the flanking genomic DNA sequence was determined for 100 to 250 nucleotides in the 5′ and 3′ directions. The length of 5′ and 3′ nucleotide sequence obtained for each rescued integration event is indicated in Fig. 5. These sequences were used to search the NCBI sequence database by using the BLAST program (1). No matches were found between these sequences and any tRNA genes. The flanking genomic DNA sequences were also analyzed with the tRNA SCAN program (25), which searches sequences for tRNA structures. The tRNA SCAN program failed to find any tRNA genes in flanking sequences. Therefore, the three independent integrations mediated by the M-MuLV V/Ty3 chimeric IN protein did not have the position specificity of Ty3.

Analysis of viral proteins and RNA. To determine the basis of low titers of G418-resistant HT1080 cells infected with the RgpA$_M$B$_M$C$_T$ chimeric retroviral vector, viral protein, and RNA levels of the RgpNeo wild-type and the RgpA$_M$B$_M$C$_T$ chimera were compared. The NC10 producer cells were used as the source of viral proteins and RNA. Western analysis of viral proteins from concentrated supernatants and cell lysates revealed high levels of mature capsid protein (30 kDa) from the RgpNeo producer cells and much (10- to 100-fold) lower levels of mature capsid protein from the RgpA$_M$B$_M$C$_T$ NC10 producer cells (data not shown). Western analysis of these same protein samples with a polyclonal anti-M-MuLV IN antibody detected protein only from the pRgpNeo producer cells, while an anti-Ty3 IN antibody could not distinguish a unique protein from any protein sample (data not shown). Similar results were obtained from immunoprecipitation of radiolabeled NC10 producer cells (data not shown). NC10 producer cell supernatants were also tested for reverse transcriptase activity. The positive control supernatants from RgpNeo-infected cells yielded significant reverse transcriptase activity, whereas the activity in chimera infected cell supernatants could not be distinguished from negative control supernatants of uninfected cells (data not shown). These results are consistent with low levels of viral proteins expressed from the NC10 chimera producer cells. Low levels of proteins could be attributable to poor expression of the chimeras or to unstable proteins.

To determine whether the low levels of viral proteins expressed in chimeric retroviral vector-producing NC10 cells were due to low levels of viral gene expression, vector RNA levels were determined. Equivalent amounts of cytoplasmic RNA (10 μg) from NC10 cells (lanes 1), RgpNeo-producing NC10 cells (lanes 2 and a lower exposure in lanes 4), and RgpA$_M$B$_M$C$_T$-producing NC10 cells (lanes 3) were subjected to Northern analysis using α-^{32}P-labeled M-MuLV *gag-pol*, M-MuLV IN, Ty3 IN, and Neo^r probes (Fig. 7). The membrane hybridized with the M-MuLV IN probe was stripped and rehybridized with the M-MuLV *gag-pol* probe and then restripped and rehybridized with the Neo^r probe. In general, the level of viral RNA expression is significantly lower from the RgpA$_M$B$_M$C$_T$ chimera-producing cells than from the RgpNeo wild-type virus-producing cells (compare lanes 2 to lanes 3). The RNAs expressed from each vector were not further inves-

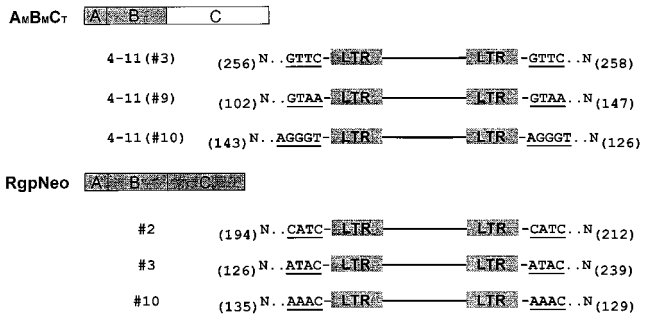

FIG. 5. Genomic target DNA sequence and identification of target site repeats. Target genomic DNA flanking the integrated retroviral vectors was subcloned into the pIBI-20 vector. Flanking genomic DNA was identified by sequence analysis with an oligonucleotide complementary to M-MuLV U5 LTR sequences (3′ target DNA) or pIBI-20 vector sequences (5′ target DNA). The three independent integrations rescued from HT1080 cells infected with either the chimeric RgpA$_M$B$_M$C$_T$ or the wild-type RgpNeo retroviral vector are illustrated. The nucleotide sequence of the genomic DNA duplicated upon integration is underlined. The lengths of flanking genomic DNA sequences obtained on the 5′ and 3′ sides of the integrated retroviral vector are indicated in parentheses. The boxes encompassing the LTRs flank the retroviral vector sequence that is represented by the solid line.

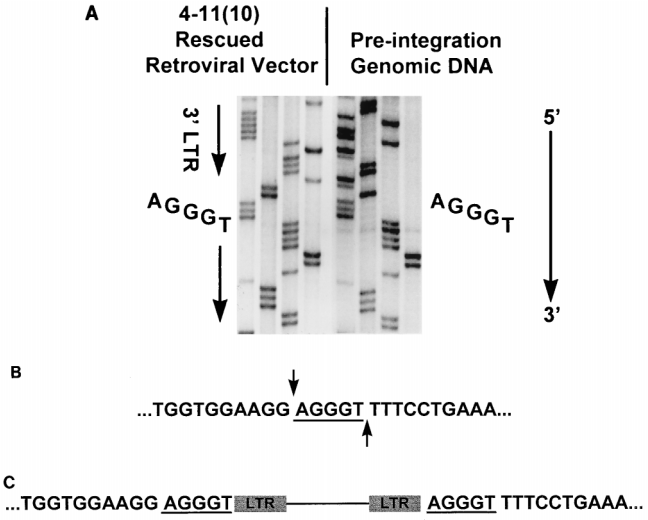

FIG. 6. Verification of the 5-bp target site repeat. (A) Preintegration genomic DNA from HT1080 cells was amplified by asymmetric PCR using oligonucleotide primers 406 and 407, complementary to the 5′ and 3′, respectively, genomic sequences flanking the rescued integration from the chimeric retroviral vector clone 4-11(10). The sequence of the PCR product was determined by using oligonucleotide 407 as a primer. The nucleotide sequence shown at the left is from the subcloned, rescued genomic DNA, representing the 3′ LTR and the adjacent flanking genomic DNA, with the 5-bp AGGGT repeat indicated. The nucleotide sequence shown at the right is from the amplified preintegration genomic DNA with one copy of the 5-bp AGGGT sequence. (B) Nucleotide sequence of the preintegration genomic DNA, with the 5-bp AGGGT sequence underlined and the positions of the staggered nicks made by the chimeric IN protein indicated by the arrows. (C) Diagram of this region after integration of the retroviral vector DNA and repair by cellular enzymes has occurred.

tigated. The longest transcript is assumed to represent the full-length vector transcript because each of the probes hybridizes to this transcript. The shortest transcript hybridizing to the Neo^r probe is assumed to be the transcript derived from the SV40 promoter upstream of the Neo^r gene. The significantly (10- to 100-fold) lower level of vector RNA expression would be sufficient to explain the low levels of viral proteins detected from the $RgpA_MB_MC_T$ chimera-producing cells.

DISCUSSION

Integration of retroviruses and retrotransposons into the host DNA displays various degrees of target site specificity. A number of factors, including the local DNA structure and the proteins bound to the DNA, influence target specificity. Specific proteins can affect insertion site selection through changes in the DNA structure or by interactions directly with the integration machinery. M-MuLV and HIV IN proteins act on the exposed major groove of DNA assembled into chromatin (54). The favored sites in this context are positions which are distorted the most by the bending of the DNA around the nucleosome (52). Targeting integration of the yeast retrovirus-like element Ty3 to tRNA genes requires the presence of pol III transcription factors on a transcriptionally competent tRNA gene template (38). This requirement suggests that integration may be targeted to tRNA genes via a protein-protein interaction between the Ty3 integration machinery and the pol III transcription factors. Although bending of the DNA in this region due to transcription factor binding occurs (44) and may enhance integration, it probably does not explain the exclusive use of tRNA genes as targets for Ty3 integration. In the work reported here, we have tested whether a retrovirus with a substitution of the C-terminal domain of M-MuLV IN with the C-terminal domain of Ty3 IN retains IN activity and whether the chimeric IN protein possess an altered target site preference. Several observations suggested that the $RgpA_MB_MC_T$ chimeric retroviral vector retained IN activity. (i) Infection of target HT1080 cells with the Neo^r-marked $RgpA_MB_MC_T$ chimeric retroviral vector yielded G418-resistant cells. (ii) The G418-resistant cells contained retrovirus insertions with intact chimeric IN-coding sequence. (iii) Each chimeric retrovirus insertion was flanked by short direct repeats, indicating that it occurred by integration. (iv) The target site repeats were 4 bp (similar to M-MuLV) and 5 bp (similar to Ty3), suggesting that the chimeric IN protein may possess a hybrid target site nicking activity. This is the first report of a chimeric IN protein retaining activity in vivo.

Ty3 IN contains the conserved sequences and 3′-end processing and strand transfer activities of retroviral IN proteins. The central core domains of M-MuLV and Ty3 IN encompassing the HHCC and DD(35)E motifs show 25% amino acid identity (4). The DD(35)E motif is absolutely required for the catalytic activity of IN since mutations in any of these residues will abolish IN activity in vitro (21, 22, 38, 40, 43, 64) and in vivo (11, 23, 39, 60, 63, 66). The conservation of functional motifs in this region suggests that it is responsible for functions of IN conserved in both Ty3 and M-MuLV IN. It therefore seems unlikely to mediate the position-specific integration activity unique to Ty3. HIV and FIV (feline immunodeficiency virus) IN display strand transfer patterns in vitro into naked DNA that are distinct from each other (59). A recent report describing the integration patterns generated in vitro by chimeric IN proteins between HIV and FIV indicated that the central core region was responsible for the distinct sequence preferences for strand transfer by these two IN proteins (59). Ty3 position-specific insertion, however, has not been demonstrated in vitro in the absence of target binding proteins and is relatively sequence independent. It therefore seems unlikely to have a distinct structural basis from the sequence preferences exhibited by HIV and FIV IN proteins.

In contrast to the central region, the N- and C-terminal domains of retroviruses are poorly conserved. Significant do-

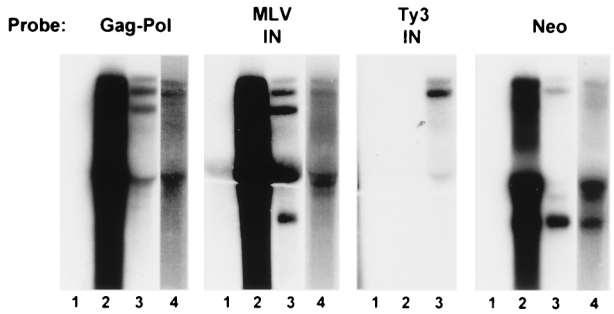

FIG. 7. Vector RNA levels in NC10 producer cells. Total cytoplasmic RNA harvested from NC10 cells (lanes 1), RgpNeo NC10 producer cells (lanes 2), and $RgpA_MB_MC_T$ NC10 producer cells (lanes 3) was denatured by glyoxylation, fractionated by electrophoresis, and transferred to a nylon membrane. Ten micrograms of RNA from each of the three samples was loaded in three separate sets on the same agarose gel. One set was stained with ethidium bromide to test for equivalent loading (data not shown). The other two sets were transferred onto nylon membranes. One set was hybridized with a Ty3 IN-coding region-specific probe, and the other set was hybridized with an M-MuLV IN-coding region-specific probe. The membrane hybridized with the M-MuLV IN probe was stripped and rehybridized with the *gag-pol*-specific probe and then stripped again and rehybridized with the Neo^r-specific probe. Lanes 1 to 3 represent the same exposure of the membranes to compare RNA levels, while lane 4 is a shorter exposure of samples in lane 2 to distinguish the hybridizing species.

main differences between retroviruses and Ty3 also exist in these regions. The domain N terminal to the HHCC motif is very small in some retroviruses and is poorly conserved among retroviral IN proteins generally. Its function is not known. This domain is 90 aa in Ty3, 55 aa in M-MuLV, and only 11 aa in HIV. Therefore, it could encode a function unique to Ty3. The C-terminal domain of retrovirus IN has been shown to have nonspecific DNA binding activity and therefore is believed to interact with the target DNA (24, 36, 49, 58, 65, 67). This domain also shows great variability among retroviral IN proteins (31, 36). The weak conservation of this domain would be consistent with the evolution of different targeting capabilities. The differences between Ty3 and retroviruses in integration patterns are extreme. However, more subtle differences are observed among retroviruses. For example, the distribution of M-MuLV IN and HIV IN insertion sites was not identical in minichromosome targets (54). In Ty3, the C-terminal domain is significantly larger than in retroviruses: 230 aa for Ty3 versus 140 aa for M-MuLV and 100 aa for HIV. The significantly larger size and sequence dissimilarity suggest that this region could mediate disparate functions in Ty3 and retroviruses. If, as the in vitro data suggest, Ty3 IN targets integration to tRNA genes via protein-protein interactions with pol III transcription factors, then the C-terminal region of Ty3 IN may be involved in this interaction.

Although the $A_MB_MC_T$ chimeric IN protein retained activity in vivo, the virus encoding this protein was much less infectious than the wild-type RgpNeo. Infection of HT1080 cells with the $RgpA_MB_MC_T$ chimeric vector resulted in the generation of only three G418-resistant HT1080 cell clones, compared to more than 10^4 cell clones for RgpNeo. There are a number of possible reasons that the $RgpA_MB_MC_T$ chimeric vector would have lower titers than the wild-type RgpNeo. These possibilities were investigated by analysis of the chimeric retroviral proteins and RNA. Hybridization with the M-MuLV IN-specific probe to the RNA from the pRgpNeo producer cells yielded bands with 10- to 100-fold-greater intensity than the bands seen with RNA from the $RgpA_MB_MC_T$ producer cells. The observed difference in RNA levels between cells expressing the wild-type RgpNeo and the $RgpA_MB_MC_T$ chimera is comparable to the difference in viral protein levels observed. The low viral protein levels seen in producer cell supernatants would explain the low titers of G418-resistant HT1080 target cells produced upon infection with these supernatants, although a reduced efficiency of integration due to additional instability or poor activity of the chimeric IN protein could also contribute. The low level of RNA expression from the $RgpA_MB_MC_T$ chimera producer cells could be explained by a genomic location or context effect on expression. However, this explanation is unlikely because the $RgpA_MB_MC_T$ chimera producer cells should represent a mixed population of chimeric proviruses integrated in different positions and therefore different genomic contexts as do the wild-type RgpNeo producer cells. In addition, wild-type M-MuLV IN mediated the integration of wild-type RgpNeo and the $RgpA_MB_MC_T$ chimeric vector into the NC10 genomic DNA at roughly the same efficiency, since similar numbers of G418-resistant NC10 cells were generated from each vector. The possibility that mutations that lowered transcription levels were introduced at some step into the LTR of the $pRgpA_MB_MC_T$ chimeric vector was investigated. The entire LTR regions of the input and rescued plasmids from the wild-type and $pRgpA_MB_MC_T$ chimeric vectors were sequenced. No sequence differences were detected between the input or the rescued plasmids from the $pRgpA_MB_MC_T$ chimeric vector and the wild-type vector. Thus, mutations in the LTR do not cause differential expression of the chimeric vector.

Substitution of the C-terminal domain of Ty3 IN for M-MuLV IN apparently provides a required function to the IN. Deletion analysis in the C-terminal domain of M-MuLV IN showed that deletions of more than 28 aa resulted in the loss of IN activity in vitro (32) and virus viability in vivo (55). Two different short linker insertions at aa 322 in M-MuLV resulted in either nonviable virus (20) or a virus with severely delayed growth (32). The ability of the Ty3 IN C-terminal domain to substitute at some level for the C-terminal domain of M-MuLV IN suggests that the function of this domain is conserved.

No gross deletions or rearrangements of the Ty3 IN sequences occurred in the three $RgpA_MB_MC_T$ chimeric virus insertions recovered. Sequence analysis verified the maintenance of the Ty3 IN sequences in this region, with the exception of single nucleotide changes identified in the 4-11(3) and 4-11(10) rescued vectors. The single nucleotide changes would result in an amino acid substitution of aspartic acid for asparagine at position 382 in 4-11(3) and lysine for glutamic acid at position 486 in 4-11(10) in the Ty3 IN C domain. The step in production of the chimeric retrovirus in which these mutations occurred would have determined whether the chimeric IN protein that actually mediated integration contained this substitution. If it was generated during reverse transcription prior to integration into NC10 cells, then the chimeric IN protein would have contained the amino acid substitution. However, if it was not generated until the reverse transcription step prior to integration in HT1080 cells, then the chimeric IN protein produced in NC10 cells would not have contained this substitution. These two possibilities could be distinguished by reverse transcription-PCR analysis of viral RNAs isolated at each stage and comparison of the nucleotide sequences in this region. If the mutation occurred at the first reverse transcription step and the chimeric protein did contain the amino acid substitution, the rescued chimeric retroviral vector plasmid could be used to generate virus and tested for IN activity.

Retrovirus and retrotransposon IN proteins make a staggered cut in the target DNA. The size of this cut is characteristic of each IN protein and determines the size of the target site repeat. The identification of both 4- and 5-bp target site repeats produced by the $RgpA_MB_MC_T$ chimeric virus suggests that the chimeric IN protein may possess a target site cutting activity which is a hybrid between M-MuLV (4 bp) and Ty3 (5 bp) IN activities. Comparison of the data from HIV and ASV IN crystal structures suggests that the distance between active-site residues in IN dimers may determine the size of the staggered cut (2). Therefore, one possible explanation for our results is that the chimeric IN protein does not form as stable a dimer as either the intact M-MuLV or Ty3 IN protein. The chimeric nature of the IN protein could affect interactions between the chimeric IN monomers or affect interactions with other viral or cellular proteins which are required in the integration complex. This less stable, chimeric IN dimer would then be less consistent in the relative positions of the staggered nicks in the target DNA. Variation in target site repeats has also been detected in products of in vitro integrations in studies using IN purified from avian myeloblastosis virions (26) or IN purified from avian sarcoma-leukosis virions or bacterial expression systems (34). In vivo, IN functions in a complex with other viral and perhaps cellular proteins; therefore, something could be missing in the in vitro reactions which contribute to the stability of the IN dimers and, in turn, the fidelity of the reaction in vivo.

Analysis of the flanking genomic DNA from three integra-

tions mediated by the $RgpA_MB_MC_T$ chimeric viruses did not reveal tRNA genes. Therefore, these integrations did not exhibit Ty3 position specificity. There are several potential explanations for the apparent lack of specificity. First, the C-terminal domain of Ty3 IN does not independently mediate the position-specific integration of Ty3. Second, specific integrations occurred, but these were not recovered efficiently. Third, this domain mediates specificity but is inactive in the chimeric context. Fourth, the C-terminal domain of Ty3 interacts with yeast target proteins but not the human homologs. It has recently been shown that Ty3 can target integration to a human tRNA gene in yeast (18). Experiments using in vitro integration assays are under way to test whether human extracts can satisfy the Ty3 integration requirement for pol III transcription factors and to determine which, if any, of several chimeric retroviral vectors have Ty3 position specificity in vitro.

Fusion of heterologous DNA binding domains to retroviral IN proteins has been successfully used in vitro to target strand transfer to regions adjacent to the binding sites (8, 10, 27, 35). The LexA DNA binding domain fused to the C terminus of ASV IN was incorporated into viral particles. However, the resulting virus displayed delayed growth kinetics (35). In this study, we have introduced an alternative novel approach to targeting integration to a specific site. A chimeric retrovirus substituting the C domain of Ty3 IN for the C domain of M-MuLV IN was generated in an attempt to confer the position-specific integration property of Ty3 on the M-MuLV retroviral vector. This strategy would exploit the existence of IN homologs that are position specific to produce predictable insertions of retrovirus vectors into preexisting genomic targets.

ACKNOWLEDGMENTS

We thank H. Fan for the gift of plasmid p2XMLV. We thank Harry Mangalam and Virginia Bilanchone for computer analysis. We thank Marielle Reyes and Dat Hoang for excellent technical assistance.

This work was supported by Public Health Service grant STTR 95-1 to Viagene, Inc., San Diego, Calif., Chiron Technologies Center for Gene Therapy, and UC-STAR grant S96-45.

REFERENCES

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
2. **Andrake, M. D., and A. M. Skalka.** 1996. Retroviral integrase, putting the pieces together. J. Biol. Chem. **271:**19633–19636.
3. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.).** 1992. Current protocols in molecular biology. Greene Publishing Associates/Wiley-Interscience, New York, N.Y.
4. **Bilanchone, V. B., and S. B. Sandmeyer.** Unpublished data.
5. **Bodner, M.** Personal communication.
6. **Bor, Y. C., F. D. Bushman, and L. E. Orgel.** 1995. In vitro integration of human immunodeficiency virus type 1 cDNA into targets containing protein-induced bends. Proc. Natl. Acad. Sci. USA **92:**10334–10338.
7. **Burns, J. C., T. Friedmann, W. Driever, M. Burrascano, and J.-K. Yee.** 1990. Vesicular stomatitis virus G glycoprotein pseudotyped retroviral vectors: concentration to a very high titer and efficient gene transfer into mammalian and nonmammalian cells. Proc. Natl. Acad. Sci. USA **90:**8033–8037.
8. **Bushman, F.** 1994. Tethering human immunodeficiency virus 1 integrase to a DNA site directs integration to nearby sequences. Proc. Natl. Acad. Sci. USA **91:**9233–9237.
9. **Bushman, F. D., A. Engelman, I. Palmer, P. Wingfield, and R. Craigie.** 1993. Domains of the integrase protein of human immunodeficiency virus type 1 responsible for polynucleotidyl transfer and zinc binding. Proc. Natl. Acad. Sci. USA **90:**3428–3432.
10. **Bushman, F. D., and M. D. Miller.** 1997. Tethering human immunodeficiency virus type 1 preintegration complexes to target DNA promotes integration at nearby sites. J. Virol. **71:**458–464.
11. **Cannon, P. M., W. Wilson, E. Byles, S. M. Kingsman, and A. J. Kingsman.** 1994. Human immunodeficiency virus type 1 integrase: effect on viral replication of mutations at highly conserved residues. J. Virol. **68:**4768–4775.
12. **Chalker, D. L., and S. B. Sandmeyer.** 1990. Transfer RNA genes are genomic targets for *de novo* transposition of the yeast retrotransposon Ty3. Genetics **126:**837–850.
13. **Chalker, D. L., and S. B. Sandmeyer.** 1992. Ty3 integrates within the region of RNA polymerase III transcription initiation. Genes Dev. **6:**117–128.
14. **Clark, D. J., V. W. Bilanchone, L. J. Haywood, S. L. Dildine, and S. B. Sandmeyer.** 1988. A yeast sigma composite element, Ty3, has properties of a retrotransposon. J. Biol. Chem. **263:**1413–1423.
15. **Connolly, C. M., and S. B. Sandmeyer.** 1997. RNA polymerase III interferes with Ty3 integration. FEBS Lett. **405:**305–311.
16. **Craigie, R.** 1992. Hotspots and warm spots: integration specificity of retroelements. Trends Genet. **8:**187–190.
17. **Dente, L., G. Cesareni, and R. Cortese.** 1983. pEMBL: a new family of single stranded plasmids. Nucleic Acids Res. **11:**1645–1655.
18. **Dildine, S. L., and S. B. Sandmeyer.** 1997. Integration of the yeast retrovirus-like element Ty3 upstream of a human tRNA gene expressed in yeast. Gene **194:**227–233.
19. **Dildine, S. L., and B. L. Semler.** 1989. The deletion of 41 proximal nucleotides reverts a poliovirus mutant containing a temperature-sensitive lesion in the 5′ noncoding region of genomic RNA. J. Virol. **63:**847–862.
20. **Donehower, L. A.** 1988. Analysis of mutant Moloney murine leukemia viruses containing linker insertion mutations in the 3′ region of *pol*. J. Virol. **62:**3958–3964.
21. **Drelich, M., R. Wilhelm, and J. Mous.** 1992. Identification of amino acid residues critical for endonuclease and integrase activities of HIV-1 IN protein in vitro. Viology **188:**459–468.
22. **Engelman, A., and R. Craigie.** 1992. Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro. J. Virol. **66:**6361–6369.
23. **Engelman, A., G. Englund, J. M. Orenstein, M. A. Martin, and R. Craigie.** 1995. Multiple effects of mutations in human immunodeficiency virus type 1 integrase on viral replication. J. Virol. **69:**2729–2736.
24. **Engelman, A., A. B. Hickman, and R. Craigie.** 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.
25. **Fichant, G. A., and C. Burks.** 1991. Identifying potential tRNA genes in genomic DNA sequences. J. Mol. Biol. **220:**659–671.
26. **Fitzgerald, M. L., A. C. Vora, W. G. Zeh, and D. P. Grandgenett.** 1992. Concerted integration of viral DNA termini by purified avian myeloblastosis virus integrase. J. Virol. **66:**6257–6263.
27. **Goulaouic, H., and S. A. Chow.** 1996. Directed integration of viral DNA mediated by fusion proteins consisting of human immunodeficiency virus type 1 integrase and *Escherichia coli* LexA protein. J. Virol. **70:**37–46.
28. **Graham, F. L., and A. J. van der Eb.** 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology **52:**456–467.
29. **Hansen, L. J., D. L. Chalker, and S. B. Sandmeyer.** 1988. Ty3, a yeast retrotransposon associated with tRNA genes, has homology to animal retroviruses. Mol. Cell. Biol. **8:**5245–5256.
30. **Hansen, L. J., and S. B. Sandmeyer.** 1990. Characterization of a transpositionally active Ty3 element and identification of the Ty3 integrase protein. J. Virol. **64:**2599–2607.
31. **Johnson, M. S., M. A. McClure, D.-F. Feng, J. Gray, and R. F. Doolittle.** 1986. Computer analysis of retroviral *pol* genes: assignment of enzymatic functions to specific sequences and homologies with non-viral enzymes. Proc. Natl. Acad. Sci. USA **83:**7648–7652.
32. **Jonsson, C. B., G. A. Donzella, E. Gaucan, C. M. Smith, and M. J. Roth.** 1996. Functional domains of Moloney murine leukemia virus integrase defined by mutation and complementation analysis. J. Virol. **70:**4585–4597.
33. **Kalpana, G. V., S. Marmon, W. Wang, G. R. Crabtree, and S. P. Goff.** 1994. Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. Science **266:**2002–2006.
34. **Katz, R. A., G. Merkel, J. Kulkosky, J. Leis, and A. M. Skalka.** 1990. The avian retroviral IN protein is both necessary and sufficient for integrative recombination in vitro. Cell **63:**87–95.
35. **Katz, R. A., G. Merkel, and A. M. Skalka.** 1996. Targeting of retroviral integrase by fusion to a heterologous DNA binding domain: in vitro activities and incorporation of a fusion protein into viral particles. Virology **217:**178–190.
36. **Khan, E., J. P. Mack, R. A. Katz, J. Kulkosky, and A. M. Skalka.** 1991. Retroviral integrase domains: DNA binding and the recognition of LTR sequences. Nucleic Acids Res. **19:**851–860.
37. **Kinsey, P. T., and S. B. Sandmeyer.** 1991. Adjacent pol II and pol III promoters: transcription of the yeast retrotransposon Ty3 and a target tRNA gene. Nucleic Acids Res. **19:**1317–1324.
38. **Kirchner, J., C. M. Connolly, and S. B. Sandmeyer.** 1995. Requirement of RNA polymerase III transcription factors for in vitro position-specific integration of a retroviruslike element. Science **267:**1488–1491.
39. **Kirchner, J., and S. B. Sandmeyer.** 1996. Ty3 integrase mutants defective in reverse transcription or 3′ end processing of extrachromosomal Ty3 DNA. J. Virol. **70:**4737–4747.
40. **Kulkosky, J., K. S. Jones, R. A. Katz, J. P. G. Mack, and A. M. Skalka.** 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:**2331–2338.
41. **Kunkel, T. A.** 1985. Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc. Natl. Acad. Sci. USA **82:**488–492.
42. **LaFemina, R. L., C. L. Schneider, H. L. Robbins, P. L. Callahan, K. LeGrow, E. Roth, W. A. Schleif, and E. A. Emini.** 1992. Requirement of active human immunodeficiency virus type 1 integrase enzyme for productive infection of human T-lymphoid cells. J. Virol. **66:**7414–7419.
43. **Leavitt, A. D., L. Shiue, and H. E. Varmus.** 1993. Site-directed mutagenesis of HIV-1 integrase demonstrates differential effects on integrase functions in vitro. J. Biol. Chem. **268:**2113–2119.
44. **Leveillard, T., G. A. Kassavetis, and E. P. Geiduschek.** 1991. *Saccharomyces cerevisiae* transcription factors IIIB and IIIC bend DNA of the $tRNA^{Gln}$ gene. J. Biol. Chem. **266:**5162–5168.
45. **McMaster, G. K., and G. G. Carmichael.** 1977. Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. Proc. Natl. Acad. Sci. USA **74:**4835–4838.
46. **Miller, A. D., and I. M. Verma.** 1984. Two base changes restore infectivity to a noninfectious molecular clone of Moloney murine leukemia virus (pMLV-1). J. Virol. **49:**214–222.
47. **Miller, M. D., and F. D. Bushman.** 1995. HIV integration. Ini1 for integration? Curr. Biol. **5:**368–370.
48. **Muller, H.-P., and H. E. Varmus.** 1994. DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes. EMBO J. **13:**4704–4714.
49. **Mumm, S. R., and D. P. Grandgenett.** 1991. Defining nucleic acid-binding properties of avian retrovirus integrase by deletion analysis. J. Virol. **65:** 1160–1167.
50. **Overhauser, J., and H. Fan.** 1996. Generation of glucocorticoid-responsive Moloney murine leukemia virus by insertion of regulatory sequences from murine mammary tumor virus into the long terminal repeat. J. Virol. **54:** 133–144.
51. **Price, J., D. Turner, and C. Cepko.** 1987. Lineage analysis in the vertebrate nervous system by retrovirus-mediated gene transfer. Proc. Natl. Acad. Sci. USA **84:**156–160.
52. **Pruss, D., F. D. Bushman, and A. P. Wolffe.** 1994. Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core. Proc. Natl. Acad. Sci. USA **91:**5913–5917.
53. **Pruss, D., R. Reeves, F. D. Bushman, and A. P. Wolffe.** 1994. The influence of DNA and nucleosome structure on integration events directed by HIV integrase. J. Biol. Chem. **269:**25031–25041.
54. **Pryciak, P. M., and H. E. Varmus.** 1992. Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. Cell **69:**769–780.
55. **Roth, M. J.** 1991. Mutational analysis of the carboxyl terminus of the Moloney murine leukemia virus integration protein. J. Virol. **65:**2141–2145.
56. **Sandmeyer, S. B., L. J. Hansen, and D. L. Chalker.** 1990. Integration specificity of retrotransposons and retroviruses. Annu. Rev. Genet. **24:**491–518.
57. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
58. **Schauer, M., and A. Billich.** 1992. The N-terminal region of HIV-1 integrase is required for integration activity, but not for DNA binding. Biochem. Biophys. Res. Commun. **185:**874–880.
59. **Shibagaki, Y., and S. A. Chow.** 1997. Central core domain of retroviral interase is responsible for target site selection. J. Biol. Chem. **272:**8361–8369.
60. **Shin, C.-G., B. Taddeo, W. A. Haseltine, and C. M. Farnet.** 1994. Genetic analysis of the human immunodeficiency virus type 1 integrase protein. J. Virol. **68:**1633–1642.
61. **Shinnick, R. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukemia virus. Nature **293:**543–548.
62. **Shoemaker, C., S. Goff, E. Gilboa, M. Paskind, S. W. Mitra, and D. Baltimore.** 1980. Structure of a cloned circular Moloney murine leukemia virus DNA molecule containing an inverted segment: implications for retrovirus integration. Proc. Natl. Acad. Sci. USA **77:**3932–3936.
63. **Taddeo, B., W. A. Haseltine, and C. M. Farnet.** 1994. Integrase mutants of human immunodeficiency virus type 1 with a specific defect in integration. J. Virol. **68:**8401–8405.
64. **van Gent, D. C., A. A. M. Oude Groeneger, and R. H. A. Plasterk.** 1992. Mutational analysis of the integrase protein of human immunodeficiency virus type 2. Proc. Natl. Acad. Sci. USA **89:**9598–9602.
65. **Vink, C., A. A. M. Oude Groeneger, and R. H. A. Plasterk.** 1993. Identification of the catalytic and DNA-binding region of the human immunodeficiency virus type 1 integrase protein. Nucleic Acids Res. **21:**1419–1425.
66. **Wiskerchen, M., and M. A. Muesing.** 1995. Human immunodeficiency virus type 1 integrase: effects of mutations on viral ability to integrate, direct viral gene expression from unintegrated viral DNA templates, and sustain viral propagation in primary cells. J. Virol. **69:**376–386.
67. **Woerner, A. M., and C. J. Marcus-Sekura.** 1993. Characterization of a DNA binding domain in the C-terminus of HIV-1 integrase by deletion mutagenesis. Nucleic Acids Res. **21:**3507–3511.