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. SANDMEYER1*

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-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 integrase 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 vitro. 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
adapted to retrovirus-based gene therapy vectors, this would result in a therapeutically active 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 polynucleotide 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 block 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 retrovirus genes containing the Ty3 C-terminal domain (Am-Bc-Cc; 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.
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 RggaABcCT and wild-type RggnNeo retroviral vectors. Approximately 10 μg of each sample of genomic DNA was digested with the restriction enzyme PstI, 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 blottedter (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 probes were synthesized by extending random primers in the presence of [α-32P]dATP with the Megaprime DNA labeling system (Amersham, Inc.). Integrated retroviral vectors were recovered by digestion of genomic DNA with extraction of DNA with phenol-chloroform, and precipitation of DNA with ethanol. SacI-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 quantify the DNA by fluorometry using a Mini TKO 100 DNA Fluorometer (Beckman 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 (5). The retroviral vector and flanking genomic sequences were subcloned by standard DNA cloning techniques (3). In general, the rescued plasmids were digested with Nhel and fragments were separated by electrophoresis in agarose gels. The Nhel fragments representing the retroviral vectors were circularized. The Nhel fragments representing the flanking genomic DNA were cloned into the Nhel site of the pBl2I-20 vector. 

**Sequence analysis.** Didexonucleotide 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'-AT GCATCTCATGACACG-3' (complementary to Ty3 nucleotides 5307 to 5332); 175, 5'-CCAGGGTGAAGTATGCTGTC-3' (complementary to Ty3 nucleotides 5042 to 5074); 387, 5'-ATGACATCTATCGGACG-3' (complementary to Ty3 nucleotides 4581 to 4596); 199, 5'-CAACGGCTGAGCTAAGTGC-3' (M-MuLV nucleotides 5317 to 5341); 387, 5'-GTCCGTTGTCTTCTGTCGAGGAGACG-3' (M-MuLV nucleotides 80 to 99); universal primer, 5'-GTAAACACGAGGCCTG TGACG-3' (complementary to pBl2I-20 nucleotides 341 to 358); and reverse primer, 5'-CA GGAGAACTGCTGACG-3' (pBl2I-20 nucleotides 202 to 219). Sequencing reactions were fractionated by electrophoresis in 8% polyacrylamide/bisacylam- lide (19:1, National Diagnostics)-8 M urea gels and visualized by autoradiog- raphy. 

**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'-CATGAGAAATACTAGGTGACTGC-3'; complementary to pIBI-20 nucleotides 341 to 358); and reverse primer, 5'-GT CTGCTGTTCCTTGGGAG-3' (complementary to Ty3 nucleotides 5307 to 5323). The 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[t101]). One microgram of HT1080 genomic DNA was used and digested with SacI and VspI. Oligonucleotide primers 406 and 407, respectively, 2 mM MgCl2, and 2.5 μl of 10H 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 (Strategene) 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 preintegration sequences (nucleotides 1906 to 2444). Membranes were stripped again and rehybridized with a fragment from the Neo' gene (HindIII to EcoRI). The probes were synthesized by extending random primers in the presence of [α-32P]dATP with the Megaprime DNA labeling system (Amersham). Membranes were hybridized to probes and washed as described in reference (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, pRgKan (Fig. 1). pRgKan contains a single LTR and gag and pol genes from M-MuLV, the Neo' gene from transposon Tn5 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 pRgNeo, derived from the BAG vector (51). Plasmid pRgNeo differs from pRgKan by having twoLTRs and sequences from the polyomavirus early region. Once pRgKan is expressed in mammalian cells, an RNA genome which is identical to pRgNeo is transcribed; therefore, viruses produced from these two retroviral vectors are identical. The retroviral vector pRgAbBaC, substituting the C domain of the Ty3 IN for the C domain of M-MuLV IN in the pRgKan 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 encapsulated 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 transfected 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 pRgNeo vector (data not shown). The resulting culture then produced retrovirus consisting of the chimeric 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 (~109 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 pRgAbBaC vector, the retroviral vector pRgAbBaC (−), 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 PstI, and subjected to Southern analysis as shown in Fig. 2. PstI cleaves within the retroviral vector on either side of the IN sequences to liberate a 6-kb (pRgKan) or 6.5-kb fragment (pRgAbBaC) (Fig. 2C). Identical Southern blots were hybridized with a M-MuLV or Ty3 IN-coding sequence specific α-32P-labeled probes (Fig. 2A and B). Genomic DNA from RggnNeo-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 RgAbBaCt-infected, G418-
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. PstI-digested genomic DNA from control HT1080 cells (lanes 2) did not hybridize with either probe. The expected size of the fragments was determined by PstI digestion of plasmids pRgpKan (lanes 6), pRgpA_pB_pC_T (lanes 7), and pRgpA_pB_pC (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. Genomic DNA was isolated from G418-resistant cells containing integrated chimeric retroviral vectors. The genomic DNA was digested with the restriction enzyme ScaI, 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
selected by kanamycin resistance. To verify the rescue of a full-length retroviral vector, the rescued plasmids were digested with NheI, 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 (pRgpAMBMCT) 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 pRgpAMBMCT 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 NheI 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 NheI 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 NheI 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 NheI. 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. BglII 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 BglII and subjected to Southern analysis. Figure 4 shows the BglII digestion pattern of input and rescued retroviral vectors pRgpNeo, 3-7, and 4-11 after hybridization with either a M-MuLV IN- or Ty3 IN-coding sequence-specific probe (Fig. 4A- B). Southern analysis of genomic DNA isolated from G418-resistant HT1080 cells infected with pRgpNeo (lanes 3), pRgpAMBMCT (lanes 4) or pRgpNeo (lanes 5). The Ty3 IN-coding sequence is represented in lanes 2. Lanes 2 to 5 contain 10 μg of high-molecular-weight genomic DNA digested with restriction enzyme PstI. PstI-digested plasmids pRgpKan (lanes 6), pRgpAMBMCT (lanes 7), and pRgpAMBMCT (lanes 8) are included as molecular weight controls. Duplicate samples were analyzed by Southern hybridization with an α-32P-labeled M-MuLV IN-specific probe (A) and an α-32P-labeled Ty3 IN-specific probe (B). Lanes 1 contain λ HindIII as a size marker. (C) Schematic diagram of PstI restriction enzyme sites on circular plasmid pRgpKan (left) and integrated pRgpNeo (right). The IN-specific probes detected a 6-kb (pRgpNeo) or a 6.5-kb (pRgpAMBMCT) PstI fragment encompassing the IN sequences.

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

<table>
<thead>
<tr>
<th>Construct</th>
<th>Isolate</th>
<th>Titer on HT1080 cells*</th>
</tr>
</thead>
<tbody>
<tr>
<td>pRgpAMBMCT</td>
<td>3-7</td>
<td>0, ≥1</td>
</tr>
<tr>
<td></td>
<td>4-11</td>
<td>2, ≥3</td>
</tr>
<tr>
<td>pRgpNeo</td>
<td>&gt;10⁶</td>
<td>&gt;10⁶</td>
</tr>
<tr>
<td>pRgpAMBMCT(−)</td>
<td>0, 0</td>
<td>0, 0</td>
</tr>
</tbody>
</table>

* G418-resistant cell colonies surviving per 10 ml of filtered, NC10 producer cell supernatant. Numbers shown represent two separate transient transfections of 293-23 cells with the retroviral vectors and the establishment of two independent NC10 producer cell lines followed by titration on HT1080 cells.

### Table 2. Rescue of Ty3 IN-containing retroviral vectors

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Construct</th>
<th>Titer on HT1080 cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>pRgpAMBMCT</td>
<td>3-7</td>
<td>0, ≥1</td>
</tr>
<tr>
<td></td>
<td>4-11</td>
<td>2, ≥3</td>
</tr>
<tr>
<td>pRgpNeo</td>
<td>&gt;10⁶</td>
<td>&gt;10⁶</td>
</tr>
<tr>
<td>pRgpAMBMCT(−)</td>
<td>0, 0</td>
<td>0, 0</td>
</tr>
</tbody>
</table>

The entire Ty3 IN C domain in the three 4-11 chimeric clones indicated that no gross deletion or rearrangements of the chimeric IN region were noted (Fig. 4D). These results were consistent with previous studies showing that the Ty3 IN C domain is functional in vivo.

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 RgpAMBMCT chimeric retroviral vector integrations was subcloned into the pIBI-20 vector (Table 2) for sequence analysis. Flanking sequence from the pRgpNeo retroviral vector, as the expected 9.2-kb retroviral vector with the wild-type RgpNeo retrovirus. Each of these contained the 9.2- or 9.7-kb 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 NheI 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 NheI. 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. BglII 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 BglII and subjected to Southern analysis. Figure 4 shows the BglII digestion pattern of input and rescued retroviral vectors pRgpNeo, 3-7, and 4-11 after hybridization with either a M-MuLV IN- or Ty3 IN-coding sequence-specific probe (Fig. 4A- B). Southern analysis of genomic DNA isolated from G418-resistant HT1080 cells infected with pRgpNeo (lanes 3), pRgpAMBMCT (lanes 4) or pRgpNeo (lanes 5). The Ty3 IN-coding sequence is represented in lanes 2. Lanes 2 to 5 contain 10 μg of high-molecular-weight genomic DNA digested with restriction enzyme PstI. PstI-digested plasmids pRgpKan (lanes 6), pRgpAMBMCT (lanes 7), and pRgpAMBMCT (lanes 8) are included as molecular weight controls. Duplicate samples were analyzed by Southern hybridization with an α-32P-labeled M-MuLV IN-specific probe (A) and an α-32P-labeled Ty3 IN-specific probe (B). Lanes 1 contain λ HindIII as a size marker. (C) Schematic diagram of PstI restriction enzyme sites on circular plasmid pRgpKan (left) and integrated pRgpNeo (right). The IN-specific probes detected a 6-kb (pRgpNeo) or a 6.5-kb (pRgpAMBMCT) PstI fragment encompassing the IN sequences.
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 RgpAMBMCT 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>
<thead>
<tr>
<th>Input chimera</th>
<th>Rescued chimera</th>
<th>Size (kb)</th>
<th>Reconstructed retroviral vector</th>
<th>Subcloned flanking sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-11</td>
<td>4-11(9)</td>
<td>9.7</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4-11(10)</td>
<td>9.7</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4-11(3)</td>
<td>9.7</td>
<td>1.0</td>
<td>5.0</td>
</tr>
<tr>
<td>3-7</td>
<td>3-7(19)</td>
<td>9.7</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>pRgpNeo</td>
<td>2</td>
<td>9.2</td>
<td>5.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>9.2</td>
<td>5.5</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>9.2</td>
<td>1.8</td>
<td></td>
</tr>
</tbody>
</table>

* Clone contained no flanking sequence.
of low titers of G418-resistant HT1080 cells infected with the 

MuLV V/Ty3 chimeric IN protein did not have the position 

fore, the three independent integrations mediated by the M-

 failed to find any tRNA genes in flanking sequences. There-

sequences for tRNA structures. The tRNA SCAN program 

alyzed with the tRNA SCAN program (25), which searches 

genes. The flanking genomic DNA sequences were also ana-

lyzed in Fig. 5. These sequences were used to search the NCBI 

sequence obtained for each rescued integration event is indi-

pRgpKan (lanes 1), pRgpAMBMCT clone 3-7 (lanes 2), and clone 4-11 (lanes 3).

Integration of the RgpAMBMCT chimeric retroviral vector is 

not adjacent to a tRNA gene. To determine whether the chimeric IN protein with the altered strand transfer activity was 

adjacent to a tRNA gene, the flanking genomic DNA sequences were identified by sequence analysis with an oligonucleotide complementary to M-MuLV LTR sequences (3’ target DNA) or pBI20 vector sequences (5’ target DNA). The three independent integrations rescued from HT1080 cells infected with either the chimeric RgpAMBMCT or the wild-type RgpNeo retroviral vector are illus-

trated. The nucleotide sequence of the genomic DNA duplicated upon integra-

tion is represented by the solid line. 

Analysis of viral proteins and RNA. To determine the basis of low titers of G418-resistant HT1080 cells infected with the 

RgpAMBMCT chimeric retroviral vector, viral protein, and 

RNA levels of the RgpNeo wild-type and the RgpAMBMCT 

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 RgpAMBMCT-NC10 

producer cells (data not shown). Western analysis of these 

same protein samples with a polyclonal anti-M-MuLV IN an-

tibody 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 radiola-

beled NC10 producer cells (data not shown). NC10 producer 

cell supernatants were also tested for reverse transcriptase 

activity. The positive control supernatants from RgpNeo-infec-

ted 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 attrib-

utable to poor expression of the chimeras or to unstable pro-

teins.

To determine whether the low levels of viral proteins ex-

pressed 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 3), and RgpAMBMCT-producing NC10 cells (lanes 3) were subjected to Northern analysis using α-32P-labeled M-MuLV gag-pol, 

M-MuLV IN, Ty3 IN, and Neo’ 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’ probe. In general, 

the level of viral RNA expression is significantly lower from the 

RgpAMBMCT 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-
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 RgpAMBC vector retained IN activity. (i) Infection of cell HT1080 cells with the Neo-r-marked RgpAMBC vector yielded G418-resistant cells. (ii) The G418-resistant cells contained retrovirus insertions with intact chimeric IN-coding sequence. (iii) Each chimeric retroviral insertion was flanked by 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 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-

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 RNA 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

FIG. 7. 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’ ends, 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 right is from the amplified preintegration genomic DNA, with the 5-bp AGGGT repeat underlined and the positions of the staggered nicks made by the chimeric IN protein indicated by the arrows. (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 region after integration of the retroviral vector DNA and repair by cellular enzymes has occurred.
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 retroviral 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 targetting 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$_p$B$_m$C$_f$-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$_m$B$_m$C$_f$-chimeric vector resulted in the generation of only one G418-resistant HT1080 cell clone, compared to more than 10$^6$ cell clones for RgpNeo. There are a number of possible reasons that the RgpA$_m$B$_m$C$_f$-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$_m$B$_m$C$_f$-producer cells. The observed difference in RNA levels between cells expressing the wild-type RgpNeo and the RgpA$_m$B$_m$C$_f$-chimeras 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$_m$B$_m$C$_f$-chimeric vector in the producer cells could be explained by a genomic location or context effect on expression. However, this explanation is unlikely because the RgpA$_m$B$_m$C$_f$-chimeric 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$_m$B$_m$C$_f$-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 RgpA$_m$B$_m$C$_f$-chimeric vector was investigated. The entire LTR regions of the input and rescued plasmids from the wild-type and RgpA$_m$B$_m$C$_f$-chimeric vectors were sequenced. No sequence differences were detected between the input or the rescued plasmids from the RgpA$_m$B$_m$C$_f$-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$_m$B$_m$C$_f$-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-hp target site repeats produced by the RgpA$_m$B$_m$C$_f$-chimeric virus suggests that the chimeric IN protein may possess a target site cutting activity which is a hybrid between M-MuLV IN (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 RgnA_BnB_Ct 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 (RNA 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 retrovector. 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 p2XML. 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 S9-45.

REFERENCES

5. Bodner, M. Personal communication.