Specific insertions of zinc finger domains into Gag-Pol yield engineered retroviral vectors with selective integration properties

Kwang-il Lim, Ryan Klimeczak, Julie H. Yu, and David V. Schaffer

Departments of Chemical Engineering and Bioengineering and The Helen Wills Neuroscience Institute, University of California, 278 Stanley Hall, Berkeley, CA 94720-3220

Edited by Mark E. Davis, California Institute of Technology, Pasadena, CA, and approved June 8, 2010 (received for review February 4, 2010)

Retroviral vectors offer benefits of efficient delivery and stable gene expression; however, their clinical use raises the concerns of insertional mutagenesis and potential oncogenesis due to genomic integration preferences in transcriptional start sites (TSS). We have shifted the integration preferences of retroviral vectors by generating a library of viral variants with a DNA-binding domain inserted at random positions throughout murine leukemia virus (Gag-Pol), then selecting for variants that are viable and exhibit altered integration properties. We found seven permissive zinc finger domain (ZFD) insertion sites throughout Gag-Pol, including within p12, reverse transcriptase, and integrase. Comprehensive genome integration analysis showed that several ZFD insertions yielded retroviral vector variants with shifted integration patterns that did not favor TSS. Furthermore, integration site analysis revealed selective integration for numerous mutants. For example, two retroviral variants with a given ZFD at appropriate positions in Gag-Pol strikingly integrated primarily into four common sites out of $3.1 \times 10^6$ possible human genome locations ($P = 4.6 \times 10^{-23}$). Our findings demonstrate that insertion of DNA-binding motifs into multiple locations in Gag-Pol can make considerable progress toward engineering safer retroviral vectors that integrate into a significantly narrowed pool of sites on human genome and overcome the preference for TSS.

The promising therapeutic success in gene therapy clinical trials for X-linked severe combined immunodeficiency has unfortunately been accompanied by the incidence of leukemia-like clonal T cell expansion in several patients, primarily arising from activation of the LMO2 protooncogene due to nearby retroviral vector integrations (1–3). While immune system function was fully rescued in the unaffected patients, the well-established preference for murine leukemia virus (MLV) integration at the start sites of transcribed regions, with the associated potential genotoxicity (2, 4, 5), represents a general risk that can offset key advantages of using these retroviruses as vectors. An alternative, lentiviruses, preferentially integrates throughout transcriptional units, rather than being concentrated near start sites (6, 7). Lentiviral infections could thus also potentially contribute to oncogenesis, though there has been no experimental evidence of this possibility to date. Various studies have suggested that viral components in the preintegration complex (PIC) in conjunction with host factors, which likely tether the complex to specific chromatin features within the host nucleus, determine retroviral and lentiviral integration patterns (8–10); however, the associated mechanisms are incompletely understood.

There have been several efforts to redirect retroviral integration via fusing sequence-specific DNA-binding domains—including the Sp1 zinc finger domain (ZFD), the DNA-binding domain (DBD) of λ phage repressor, and an engineered ZFD—to the C or N terminus of retroviral integrase (11–14), a critical determinant of integration patterns. The resulting integration behavior was monitored in vitro (11, 12) or in vivo (13, 14) using agarose-based and PCR-based assays. However, likely due to the need to coin incorporate wild-type Gag-Pol polypeptide to compensate for viral infectivity completely deprived by the engineered integrase fusions, as well as potential off-target binding of DNA-binding motifs, only modest increases in integration at the intended target site were observed.

In this study we attempted to develop safer retroviral vector systems with high infectivity that do not favor transcriptional start sites (TSS) for integration via inserting an engineered DBD into numerous permissive locations identified in MLV Gag-Pol. Given the incomplete knowledge of the composition of the PIC, and the regions within Gag-Pol that steer integration directly or by association with host factors, the optimal insertion sites for an exogenous DBD to direct integration and/or disrupt viral domains that contribute to wild-type integration preferences is not clear. Accordingly, in this study we have applied a high-throughput protein engineering approach by generating a library of viruses with DBDs inserted into random locations throughout Gag and Pol, without coin incorporation of wild-type Gag-Pol polypeptide, and selecting for variants that are viable and avoid integration into TSS. Engineered zinc finger domains (ZFDs) were chosen as the DBD for the modular binding properties of their zinc finger subunits, which enables the engineering of ZFDs with selectivity for a number of DNA sequences (15–17), as well as for their considerable albeit imperfect selectivity for such target sequences (18, 19). Our genome-wide analysis indicates that when inserted into key regions of Gag-Pol, such DBDs can override the intrinsic properties of MLV vectors to shift integration patterns toward safer regions of the genome that lack TSS.

Results and Discussion

Library Construction and Selection Results in Numerous Viable MLV Variants with ZFD Insertions in Gag and Pol. We first constructed a large ($4.3 \times 10^5$) retroviral library where a 186 amino acid polydactyl zinc finger domain ZFD1—a six zinc finger domain previously designed to recognize an 18-bp sequence (each finger binds a 3-bp sequence) that appears proximal to the γ-globin locus in the human genome (15)—was randomly inserted through the use of a transposon system (20) into likely every position of the MLV Gag and Pol proteins (Figs. 1 and 2 and Fig. S14). ZFD sequence optimization and a low copy number plasmid were required to avoid plasmid recombination issues. We then selected the library for Gag-Pol.ZF mutant clones that could package in

Author contributions: K.L., R.K., and D.V.S. designed research; K.L. performed research; J.H.Y. contributed new reagents; K.L. and D.V.S. analyzed data; and K.L. and D.V.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

1 To whom correspondence may be addressed: Department of Chemical Engineering, University of California, 176 Stanley Hall, Berkeley, CA 94720-3220. E-mail: schaffer@berkeley.edu.

This article contains supporting information online at www.pnas.orglookup/suppl/doi:10.1073/pnas.1001402107—1/DCSupplemental.

 bers. For example, the ZFD insertions for 16.p12.zfd1 and 52.p12.zfd1—approximately 2,000 copies of each Gag protein (and considerably protein stoichiometry within retroviral particles is not known, but ap-
positions 19 and 8, respectively, T able 1) but distant from the
was
benefit of integration patterns shifted away from TSS, and further
These reduced titers could conceivably be offset by the major
organization of MLV
The 14.RT.zfd1 and 375.RT.zfd1 clones harbored insertions—
within the structured αc chain of the RT fingers domain and within
the unstructured region adjacent to the start of the short αc chain of the palm domain (Fig. S3) (23), respectively. Both the finger and palm domains are known to be critical for reverse transcription via their interactions with the primer-template duplex (9), but the residues adjacent to the insertions are not involved in direct DNA contacts (Fig. S3) (24).
Three clones—15.IN.zfd1, 17.IN.zfd1, and 273.IN.zfd1—
tained the ZFD next to the second putative α helix near
the C terminus of IN, within the first putative β-sheet of the N
terminus of IN, and within the sixth α-helix of the catalytic core
domain, respectively (25). All three insertions were near IN
regions that are relatively variable among eight different MLV
strains, while no insertions were observed within the three most
conserved domains (residues 254–270, 346–362, and 368–375)
(25). In addition, 17.IN.zfd1 harbored the ZFD 19 amino acids
upstream of the HHCC domain that binds to a zinc ion, which
stabilizes the N-terminal domain and is important for IN
multimerization. The functional form of IN has been suggested to be a
multimer (perhaps a tetramer or even higher order) (9), and it is
thus surprising that the ZFD insertion near the HHCC domain
still permitted packaging of infectious particles. Finally, the
C-terminal region of IN is known to have nonspecific DNA-bind-
properties (9), and the incorporation of targeting ZFD into
the region may thus modulate the integration specificity.

Structural Context for Permissive ZFD Insertion Sites. The exact pro-
tein stoichiometry within retroviral particles is not known, but ap-
approximately 2,000 copies of each Gag protein (and considerably fewer of each Pol protein) are known to be present in a virion (9), such that the ZFDs in some cases are present in high copy num-
bers. For example, the ZFD insertions for 16.p12.zfd1 and 52.
p12.zfd1—within the early region of p12 (starting at amino acid
positions 19 and 8, respectively, Table 1) but distant from the
PNP motif (residues 31–34) that is critical for efficient virus re-
lease from the cells (21)—would interestingly be present at very
high copy numbers per virion. Because p12 is composed only of
random coils (22), this unstructured protein may have sufficient
flexibility to accommodate large protein domains such as the
ZFD. Based on the roles of p12 in integration (22), the ZFD-
p12 fusion presumably remains associated with the PIC and is
thus ultimately in a position to affect viral integration patterns.
The 14.RT.zfd1 and 375.RT.zfd1 clones harbored insertions—
within the structured αc chain of the RT fingers domain and within
the unstructured region adjacent to the start of the short αc chain of the palm domain (Fig. S3) (23), respectively. Both the finger and palm domains are known to be critical for reverse transcription via their interactions with the primer-template duplex (9), but the residues adjacent to the insertions are not involved in direct DNA contacts (Fig. S3) (24).

Three clones—15.IN.zfd1, 17.IN.zfd1, and 273.IN.zfd1—
tained the ZFD next to the second putative α helix near
the C terminus of IN, within the first putative β-sheet of the N
terminus of IN, and within the sixth α-helix of the catalytic core
domain, respectively (25). All three insertions were near IN
regions that are relatively variable among eight different MLV
strains, while no insertions were observed within the three most
conserved domains (residues 254–270, 346–362, and 368–375)
(25). In addition, 17.IN.zfd1 harbored the ZFD 19 amino acids
upstream of the HHCC domain that binds to a zinc ion, which
stabilizes the N-terminal domain and is important for IN
multimerization. The functional form of IN has been suggested to be a
multimer (perhaps a tetramer or even higher order) (9), and it is
thus surprising that the ZFD insertion near the HHCC domain
still permitted packaging of infectious particles. Finally, the
C-terminal region of IN is known to have nonspecific DNA-bind-
properties (9), and the incorporation of targeting ZFD into
the region may thus modulate the integration specificity.

ZFD Insertion Variant Did Not Favor Transcriptional Start Sites for Integra-
tion. We next adapted a high-throughput method to analyze
genomic integration patterns of wild-type and mutant MLV
vectors, originally developed for analysis of HIV-1 integration
(SI Text). This method, whose results were previously validated
by statistical comparison to conventional integration analysis
(26), uses the type IIS restriction enzyme Mme I to generate short
fragments containing virus-host genome junctions to enable effi-
cient analysis of viral integration patterns. We analyzed a total of
809 sequenced virus-host genome junctions for wild-type

Fig. 2. Schematic of library selection and clonal analysis. (i) Packaging of library viruses via transfection of plasmid library and VSV-G helper plasmid into 293T
cells. (ii) Harvest and purification of viruses. (iii) Infection of 293T cells with virion library at a multiplicity of infection (moi) of <0.1. (iv) Sorting and subsequent
expansion of GFP-positive cells. (v) Rescue of packageable mutants by 293T cell transfection with helper plasmid encoding the VSV-G envelope protein.
(vi) Subcloning Gag-Pol.ZF cDNA from packageable mutants into pCMV Gag-Pol helper plasmid. (vii) Packaging virus clones via transfection of 293T cells with
pCLPIT GFP and helper plasmids. (viii) Analysis of infectivity and genome-wide integration analysis.
Table 1. Permissive ZFD insertion sites within MLV Gag and Pol proteins

<table>
<thead>
<tr>
<th>Clones</th>
<th>Sequence near inserted zinc fingers</th>
<th>Region of insertion</th>
</tr>
</thead>
<tbody>
<tr>
<td>16.p12.zfd1</td>
<td>VLSD&lt;sup&gt;a&lt;/sup&gt;SAZF_domainAGPH5DSGGP</td>
<td>p12</td>
</tr>
<tr>
<td>16.p12.zfd2</td>
<td>VLSD&lt;sup&gt;a&lt;/sup&gt;SAZF_domainAGPH5DSGGP</td>
<td>p12</td>
</tr>
<tr>
<td>16.p12.zfd3</td>
<td>VLSD&lt;sup&gt;a&lt;/sup&gt;SAZF_domainAGPH5DSGGP</td>
<td>p12</td>
</tr>
<tr>
<td>52.p12.zfd1</td>
<td>TPSSL&lt;sup&gt;a&lt;/sup&gt;VAZF_domainAGPH5LSGAK</td>
<td>p12</td>
</tr>
<tr>
<td>52.p12.zfd2</td>
<td>TPSSL&lt;sup&gt;a&lt;/sup&gt;VAZF_domainAGPH5LSGAK</td>
<td>p12</td>
</tr>
<tr>
<td>52.p12.zfd3</td>
<td>TPSSL&lt;sup&gt;a&lt;/sup&gt;VAZF_domainAGPH5LSGAK</td>
<td>p12</td>
</tr>
<tr>
<td>14.RT.zfd1</td>
<td>EARL&lt;sup&gt;a&lt;/sup&gt;SAZF_domainAGPH5GK</td>
<td>RT</td>
</tr>
<tr>
<td>14.RT.zfd2</td>
<td>EARL&lt;sup&gt;a&lt;/sup&gt;SAZF_domainAGPH5GK</td>
<td>RT</td>
</tr>
<tr>
<td>14.RT.zfd3</td>
<td>EARL&lt;sup&gt;a&lt;/sup&gt;SAZF_domainAGPH5GK</td>
<td>RT</td>
</tr>
<tr>
<td>375.RT.zfd1</td>
<td>DVSLL&lt;sup&gt;a&lt;/sup&gt;VZPZ_domainAGPH5GK</td>
<td>RT</td>
</tr>
<tr>
<td>375.RT.zfd2</td>
<td>DVSLL&lt;sup&gt;a&lt;/sup&gt;VZPZ_domainAGPH5GK</td>
<td>RT</td>
</tr>
<tr>
<td>375.RT.zfd3</td>
<td>DVSLL&lt;sup&gt;a&lt;/sup&gt;VZPZ_domainAGPH5GK</td>
<td>RT</td>
</tr>
<tr>
<td>15.IN.zfd1</td>
<td>HVKA&lt;sup&gt;a&lt;/sup&gt;VRZF_domainAGPH5GK</td>
<td>IN</td>
</tr>
<tr>
<td>15.IN.zfd2</td>
<td>HVKA&lt;sup&gt;a&lt;/sup&gt;VRZF_domainAGPH5GK</td>
<td>IN</td>
</tr>
<tr>
<td>15.IN.zfd3</td>
<td>HVKA&lt;sup&gt;a&lt;/sup&gt;VRZF_domainAGPH5GK</td>
<td>IN</td>
</tr>
<tr>
<td>17.IN.zfd1</td>
<td>KYWV&lt;sup&gt;a&lt;/sup&gt;SAZF_domainAGPH5GK</td>
<td>IN</td>
</tr>
<tr>
<td>17.IN.zfd2</td>
<td>KYWV&lt;sup&gt;a&lt;/sup&gt;SAZF_domainAGPH5GK</td>
<td>IN</td>
</tr>
<tr>
<td>17.IN.zfd3</td>
<td>KYWV&lt;sup&gt;a&lt;/sup&gt;SAZF_domainAGPH5GK</td>
<td>IN</td>
</tr>
<tr>
<td>273.IN.zfd1</td>
<td>GAPPP&lt;sup&gt;a&lt;/sup&gt;RPZPZ_domainAGPH5GK</td>
<td>IN</td>
</tr>
<tr>
<td>273.IN.zfd2</td>
<td>GAPPP&lt;sup&gt;a&lt;/sup&gt;RPZPZ_domainAGPH5GK</td>
<td>IN</td>
</tr>
<tr>
<td>273.IN.zfd3</td>
<td>GAPPP&lt;sup&gt;a&lt;/sup&gt;RPZPZ_domainAGPH5GK</td>
<td>IN</td>
</tr>
</tbody>
</table>

The clone number is followed by the viral protein into which the ZFD was inserted, and the identity of the ZFD. The notations zfd1, zfd2, and zfd3 indicate clones containing a ZFD originally engineered to bind to a 18-bp stretch that appears in the promoter regions of γ-globin, PDEB, and CHK2 genes, respectively (15–17). The positions of amino acid residues flanking the ZFD insertion are also indicated, with the residues of wild-type MLV Gag-Pol proteins, and amino acids additionally introduced via the transposon reaction and cloning steps (underlined).

and mutant vectors in this study (Table S1). Consistent with prior reports (5), MLV vector with wild-type Gag-Pol exhibited a strong preference in the human embryonic kidney 293 cell genome for TSS and CpG islands that are known to be enriched in promoters (27) (P = 1.392 × 10<sup>−39</sup> and P = 4.926 × 10<sup>−151</sup> compared to the hypothetical case of random integration, respectively) (Table 2). In stark contrast, none of the 76 sequenced junctions for a representative clone 17.IN.zfd1 (Table S1) was located within 5 kb of a TSS (Table 2), a statistically significant shift from wild-type MLV (P = 0.0214). This mutant also showed integration frequencies in CpG islands that were intermediate between wild-type virus and random integration, but were only marginally significantly distinct from either (P = 0.0817 and P = 0.0660 for comparison to wild-type and random, respectively) (Table 2). In addition, while vector with wild-type Gag-Pol disfavored repeat elements for integration (25.2% vs. 44.6% for hypothetical random integration), the mutant did not disfavor these elements (50.0%).

![Fig. 3](image-url) **Fig. 3.** Genomic (A) and infectious (B) titers of wild-type virus and ZFD insertion mutants. (A) A standard retrorviral vector that expresses the tetra-cycline-Off transcription factor, and drives GFP expression from a tetracycline-responsive promoter (Fig. 5A), was packaged using pCMV Gag-Pol helper plasmids containing wild-type and ZFD variant Gag-Pol vectors. Vector genomic titers and standard errors of the mean are shown. (B) Infectious titers of viruses were measured via duplicate infections of 293T cells, and standard deviations of infectious titer are shown by bars.
(Table 2). For several other ZFD1 mutants, low numbers of sequenced virus-host genome junctions (Table S1) precluded statistical analysis.

Overall, these data indicate that a DBD insertion into a key location of Gag-Pol can significantly shift the overall integration patterns of MLV toward potentially safer regions where TSS are relatively rare. That said, the identified integration locations for the mutant were not near the two addresses (5,271,202 and 5,276,126 on chromosome 11 of human genome) that the ZFD1 was originally designed to target (15), which may be consistent with the potential for off-target binding of ZFDs (18, 19). Regardless, the leukemia cases observed in the X-linked severe combined immunodeficiency trial were attributed to retroviral integrations 3 kb upstream and 2 kb downstream of the LMO2 protooncogene (2); therefore, a significant reduction in the integration preference for TSS as shown in the case of 17.IN.zfd1 likely yields a safer vector.

Two ZFD Variants Integrated with High Frequency into the Same Location in the Human Genome. In addition to the global changes in the genomic integration patterns for 17.IN.zfd1, analysis of specific integrations intriguingly revealed that variants with ZFD1 inserted into different sites of Gag-Pol mediated integration into the same location in the human genome. That is, 83% and 91% of sequenced junctions for 16.p12.zfd1 and 375.RT.zfd1, respectively, strikingly mapped to the same location, nucleotide 141,457,970 on chromosome 3 (Table 3). Such a rare coincidence of integrations at the same site out of $1.6 \times 10^9$, $3.1 \times 10^9$ possible human genome locations revealed that insertion of polydactyl ZFDs into the appropriate locations in Gag-Pol can narrow the range of retroviral integrations dramatically (P = 2.6 $\times$ 10$^{-9}$, SI Materials and Methods). In strong contrast, none of the 294 sequenced virus-host genome junctions from three independent infections with MLV vector containing wild-type Gag-Pol (Table S1) mapped to this site on chromosome 3.

Identified Permissive ZFD Insertion Sites Tolerated Other ZFDs of Similar Length. To analyze whether the Gag-Pol sites that permitted insertion of the ZFD1 could also tolerate other ZFDs, and whether new ZFDs in these positions could shift integration to different locations in a modular fashion, we inserted ZFDs previously designed to bind to 18-bp sequences that appear in the pigment epithelium-derived factor (PEDF) or CHK2 checkpoint homolog (CHK2) gene promoters (16, 17) (ZFD2 and ZFD3, respectively) (Table 1). We first packaged eight mutants with ZFD2 or ZFD3 grafted into the insertion positions of clones 17, 52, 273, and 375 (Table 1), which previously showed the highest infectivities (Fig. 3B). All of the new Gag-Pol.ZF clones allowed the production of infectious virus particles (Fig. S4), importantly indicating that the insertion sites discovered within Gag-Pol proteins could be generally permissive to other polydactyl ZFDs of similar length (~180 amino acids).

Overall, these data indicate that a DBD insertion into a key location of Gag-Pol can significantly shift the overall integration patterns of MLV toward potentially safer regions where TSS are relatively rare. That said, the identified integration locations for the mutant were not near the two addresses (5,271,202 and 5,276,126 on chromosome 11 of human genome) that the ZFD1 was originally designed to target (15), which may be consistent with the potential for off-target binding of ZFDs (18, 19). Regardless, the leukemia cases observed in the X-linked severe combined immunodeficiency trial were attributed to retroviral integrations 3 kb upstream and 2 kb downstream of the LMO2 protooncogene (2); therefore, a significant reduction in the integration preference for TSS as shown in the case of 17.IN.zfd1 likely yields a safer vector.

<table>
<thead>
<tr>
<th>Clones</th>
<th>Virus-host junction sequence</th>
<th>The number of junctions with the shown sequence</th>
<th>Total number of sequenced junctions for each mutant</th>
<th>Percent out of the total number of sequenced junctions</th>
<th>Integration position on the human genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>16.p12.zfd1</td>
<td>ACACAACGTCTTAAATCAATCCA</td>
<td>5</td>
<td>6</td>
<td>83.3</td>
<td>3</td>
</tr>
<tr>
<td>375.RT.zfd1</td>
<td>ACACAACGTCTTAAATCTTTAG</td>
<td>31</td>
<td>34</td>
<td>91.2</td>
<td></td>
</tr>
<tr>
<td>273.IN.zfd2</td>
<td>ACACATTTGGGGGCCTGGACCACT</td>
<td>8</td>
<td>197</td>
<td>4.1</td>
<td>2</td>
</tr>
<tr>
<td>375.RT.zfd2</td>
<td>ACACATTTGGGGGCCTGGACCACT</td>
<td>5</td>
<td>93</td>
<td>5.4</td>
<td></td>
</tr>
<tr>
<td>273.IN.zfd2</td>
<td>ACACATTTGGGGGCCTGGACCACT</td>
<td>22</td>
<td>197</td>
<td>11.2</td>
<td>18</td>
</tr>
<tr>
<td>375.RT.zfd2</td>
<td>ACACATTTGGGGGCCTGGACCACT</td>
<td>7</td>
<td>93</td>
<td>7.5</td>
<td></td>
</tr>
<tr>
<td>273.IN.zfd2</td>
<td>ACACATTTGGGGGCCTGGACCACT</td>
<td>1</td>
<td>197</td>
<td>0.5</td>
<td>15</td>
</tr>
<tr>
<td>375.RT.zfd2</td>
<td>ACACATTTGGGGGCCTGGACCACT</td>
<td>2</td>
<td>93</td>
<td>2.2</td>
<td></td>
</tr>
<tr>
<td>273.IN.zfd2</td>
<td>ACACATTTGGGGGCCTGGACCACT</td>
<td>47</td>
<td>197</td>
<td>23.9</td>
<td>1</td>
</tr>
<tr>
<td>375.RT.zfd2</td>
<td>ACACATTTGGGGGCCTGGACCACT</td>
<td>29</td>
<td>93</td>
<td>31.2</td>
<td>9</td>
</tr>
</tbody>
</table>

The common junction sequences found in mutants with the same ZFD inserted into different Gag-Pol positions are shown, along with the fraction of total junctions corresponding to that integration position for each mutant. While most junction sequences matched a single location in the human genome, the junction sequence in the bottom two rows matched two locations. The underlined sequences are viral, and the others are from the host genome. During 3’ processing after reverse transcription the viral dinucleotides TT are deleted (9), producing a 5’ overhang that is often but not always filled post integration.
chromatin may further modulate affinities for DNA targets. Consistent with these observations, our retroviral vectors succeeded in fundamentally shifting integration patterns; however, as with the ZFD1, the new ZFDs did not find the addresses that the DBDs were originally designed to target.

That said, sequence analysis of individual virus-host genome junctions for new ZFDs strikingly revealed that the two mutants with ZFD2 in IN and RT integrated into a small set of common locations in the human genome. That is, 48% (of 93) and 40% (of 197) analyzed junctions for 375.RT.zfd2 and 273.IN.zfd2 (Table S1), respectively, intriguingly corresponded to a set of five common integration junction sequences despite the ZFD insertion into different viral proteins (four of these five sequences corresponded to unique positions in the human genome, Table 3). The chance that random integration could account for this high degree of coincidence in integration events at these common sites is statistically improbable (P = 4.6 × 10−29, SI Materials and Methods). However, infections of 293 cells with 273.IN.zfd2 and 273.IN.zfd3 did not result in integrations into common sites (Table S2) indicating that different ZFDs at the same site within Gag-Pol direct integrations into different sites of the genome. These observations further illustrate that insertion of ZFDs into RT and IN dramatically shrinks the pool of sites on human genome for the retroviral integrations, thereby potentially creating safer vectors.

Conclusions
In this study, we demonstrated that insertion of DBDs into key sites within Gag-Pol can engineer likely safer retroviral vectors by modulating and shifting integration patterns toward regions where TSS are relatively rare, as well as in many cases considerably narrowing the range of integration positions in the genome. This high-throughput engineering approach can also be extended to engineer other retroviruses, including lentiviruses. Therefore, in general library-based protein engineering of vectors to modulate their integration or other processes is a powerful approach to enhance the properties of viruses for clinical application.

Materials and Methods
Cell Lines and Plasmids. Human embryonic kidney 293T cells were cultured in Iscove’s modified Dulbecco’s medium with 10% fetal bovine serum at 37°C and 5% CO₂. The retroviral vector plasmids pCLGIT Gag-Pol.ZF and pCLPIT GFP (Fig. S1A) were used for virus packaging during library production and virus clonal analysis, respectively. Three point mutations were made to introduce a Mme I site into the US region of these retroviral vector plasmids (Fig. S1B), which enabled adaptation of a prior method (26) to identify the sites of retroviral integration into human genome using a high-throughput linear amplification mediated PCR. The helper plasmids pCMV Gag-Pol and pDNA3 I SVS VSV-G express MLV gag-pol and the vesicular stomatitis virus glycoprotein (VSV-G), respectively, from the cytomegalovirus (CMV) immediate-early promoter. To avoid recombination problems, the ColE1 replication origin for pCMV Gag-Pol was replaced with a low copy number immediate-early promoter. For the retroviral integrations, thereby potentially creating safer vectors.

Clonal Analysis. Each virus clone was packaged by transient transfection of pCMV Gag-Pol.ZF, pDNA3 I SVS VSV-G, and pCLPIT GFP plasmids into 293T cells. MLV supernatant was used for infection either just after filtration with 0.45-μm syringe filters or after concentration via ultracentrifugation. The infectivity titers of viruses were measured by real-time qPCR (35) using the iCyCler iQ Real Time Detection System (Bio-Rad) and SYBR Green I (Invitrogen) with primers 5′- ATTGACTGAGTCGCCCGG-3′ (forward) and 5′- ACGGAGACCCACTGCGAT-3′ (reverse). The infected cells were analyzed in six serial log dilutions via qPCR. Infectious titers of viruses were measured in duplicate by counting GFP-positive, infected cells via flow cytometry on an EPICS XL-MCL cytometer (Beckman-Coulter).

ACKNOWLEDGMENTS. We thank J. T. Koerber and Sung-kuk Lee for technical advice. This work was supported by National Science Foundation BES-0629202 and National Institutes of Health R01 GM73058.