JOURNAL OF VIROLOGY, Feb. 2006, p. 1939–1948 Vol. 80, No. 4
0022-538X/06/$08.00+0 doi:10.1128/JVI.80.4.1939–1948.2006

Human Immunodeficiency Virus Type 1 Incorporated with Fusion Proteins Consisting of Integrase and the Designed Polydactyl Zinc Finger Protein E2C Can Bias Integration of Viral DNA into a Predetermined Chromosomal Region in Human Cells

Wenjie Tan,[1] Zheng Dong,[1] Thomas A. Wilkinson,[1] Carlos F. Barbas III,[2] and Samson A. Chow[1]*

Department of Molecular and Medical Pharmacology, Molecular Biology Institute, and UCLA AIDS Institute, UCLA School of Medicine, Los Angeles, California 90095,[1] and The Skaggs Institute for Chemical Biology and the Department of Molecular Biology, The Scripps Research Institute, La Jolla, California 92037[2]

Received 26 October 2005/Accepted 1 December 2005

In vitro studies using fusion proteins consisting of human immunodeficiency virus type 1 integrase (IN) and a synthetic polydactyl zinc finger protein E2C, a sequence-specific DNA-binding protein, showed that integration of retroviral DNA can be biased towards a contiguous 18-bp E2C-recognition site. To determine whether the fusion protein strategy can achieve site-specific integration in vivo, viruses were prepared by cotransfection and various IN-E2C fusion proteins were packaged in *trans* into virions. The resulting viruses incorporated with the IN-E2C fusion proteins were functional and capable of performing integration at a level ranging from 1 to 24% of that of viruses containing wild-type (WT) IN. Two of the more infectious viruses, which contained E2C fused to either the N (E2C/IN) or to the C (IN/E2C) terminus of IN, were tested for their ability to direct integration into a unique E2C-binding site present within the 5′ untranslated region of *erbB-2* gene on human chromosome 17. The copy number of proviral DNA was measured using a quantitative real-time nested-PCR assay, and the specificity of directed integration was determined by comparing the number of proviruses within the vicinity of the E2C-binding site to that in the whole genome. Viruses containing IN/E2C fusion proteins had sevenfold higher preference for integrating near the E2C-binding site than those viruses containing WT IN, whereas viruses containing E2C/IN had 10-fold higher preference. The results indicated that the IN-E2C fusion protein strategy is capable of directing integration of retroviral DNA into a predetermined chromosomal region in the human genome.

Integration of retroviral DNA is catalyzed by the viral enzyme integrase (IN) (for reviews, see references 6 and 32). The process is nonspecific but not random (15, 25). The frequency of use of specific sites varies considerably, with some sites being preferred up to several hundred times more than random (54, 61, 62). Because of its nonspecific nature, integration of retroviral DNA is inherently a mutagenic event and can deregulate cellular gene expression by insertional mutagenesis. Studies on murine mammary tumor virus and other slow-acting retroviruses have established that integration can cause tumors either by activating proto-oncogenes or by producing a new chimeric protein encoded by the truncated, affected cellular gene and the proviral genome (43, 46). For human immunodeficiency virus type 1 (HIV-1), the incidence of human neoplasms caused by insertional mutagenesis is rare, although a common integration site for HIV-1 DNA is found near the *c-fps/fes* oncogene in AIDS-associated T-cell lymphoma (56). In addition, a variety of chimeric mRNAs were detected in HIV-1-infected H9 cells, in peripheral blood mononuclear cells, and cells in bronchoalveolar washes of AIDS patients, indicating that HIV-1 can activate transcription of cellular genes by promoter insertion (50). Due to the efficacy of the combination chemotherapy, host pathogenicity resulting from insertional mutagenesis may become more evident as the life expectancy of infected individuals lengthens.

Retroviruses are commonly engineered for use as gene delivery vehicles in mammalian cells. The retrovirus-based vectors developed to date exhibit the same salient feature as natural retroviruses of nonspecific integration throughout the genome (44, 54, 62). The ability to direct integration into a specific chromosomal region will have a major impact on the use and safety of retrovirus-based vectors in genetic engineering and therapy (21, 58, 59). The risk of insertional mutagenesis is further highlighted by the report of leukemia in mice and infants after the administration of retrovirus-based vectors in gene therapy protocols (24, 38). Since the process of integration site selection is poorly understood, we and others have tested in vitro the feasibility of directing integration into specific DNA sites by use of a fusion protein composed of a full-length HIV-1 or avian sarcoma virus IN and the *Escherichia coli* LexA repressor or the DNA-binding domain of λ repressor (9, 22, 31). The fusion proteins direct integration by recognizing and binding to their cognate target sites on the DNA, causing integration to be mediated into the immediately adjacent regions (for reviews, see references 8 and 53). In addition to purified proteins, preintegration complexes containing wild-type IN and IN-Zif268 fusion proteins can direct integration near the Zif268-recognition site (10).

A major limitation of this strategy is that the DNA-binding

* Corresponding author. Mailing address: Department of Molecular and Medical Pharmacology, Molecular Biology Institute, and UCLA AIDS Institute, UCLA School of Medicine, Los Angeles, CA 90095. Phone: (310) 825-9600. Fax: (310) 825-6267. E-mail: schow@mednet.ucla.edu.

sequences of the previously tested fusion proteins are defined and fixed and may not present at the desired chromosomal site for integration. In addition, these DNA-binding proteins can recognize multiple DNA sequences related to their consensus binding sequence or the number of nucleotides required for specific protein-DNA interaction is insufficient for specifying a unique site within a mammalian genome (29, 37, 48). One class of DNA-binding proteins that offers several advantages in conferring site specificity to retroviral IN is the synthetic polydactyl proteins derived from the Cys_2-His_2 zinc finger proteins (for reviews, see references 2 and 5). One example of such a protein is the six-zinc finger E2C, constructed by grafting the amino acid residues of each zinc finger involved in specific DNA recognition into the framework of the designed consensus protein Sp1C, a derivative of Sp1 (17). E2C binds with high affinity (in subnanomolar range) and recognizes a unique, contiguous 18-bp sequence in the human genome (3, 4). In addition, the polydactyl zinc-finger protein has a modular organization, and the fingers may be "mixed and matched" to provide proteins with new DNA-binding specificities (4, 14, 41, 51, 55). Therefore, it may be possible to design a zinc finger protein with specificity for any desired sequence. The application of this class of designer DNA-binding proteins has been illustrated by studies in which artificial transcription factors based on modified zinc-finger domains are used to activate or repress expression of reporter genes, as well as endogenous genes in the native chromosomal environment of animal and plant cells (4, 13, 18, 23, 30, 33, 40, 52). We have constructed several fusion proteins consisting of HIV-1 IN and E2C. The fusion proteins are catalytically active and bias integration of retroviral DNA near the E2C-binding site in vitro (57). The results demonstrate that the IN-E2C fusion proteins offer an efficient approach and a versatile framework for directing the integration of retroviral DNA into a specific DNA site.

At present, it is not known whether any of the aforementioned fusion proteins are able to direct integration in vivo (10, 31). A major problem in measuring the level of site specificity in mammalian cells is the existence of numerous potential binding sites for the previously described fusion proteins. With the IN-E2C fusion proteins, the E2C-recognition sequence is located within the 5′ untranslated region of the *erbB-2* gene (4). A BLAST search of the GenBank database (human 2004 freeze) confirmed that the *e2c* site is unique to *erbB-2* (chr. 17q12). We have used an in *trans* method (20, 27, 64) to incorporate the IN-E2C fusion proteins into infectious virions and subsequently evaluated the ability of the modified virus to integrate the viral DNA into a predetermined chromosomal region in cultured human cells by fluorescence-monitored real-time nested PCR.

MATERIALS AND METHODS

Cells and antibodies. HeLa (CCL-121) and 293T cells were obtained from the American Type Culture Collection. The cells were grown in Dulbecco's modified Eagle's medium (DMEM; Gibco-BRL) with 10% fetal calf serum (HyClone), 100 U of penicillin (Cellgro) per ml, and 0.1 mg of streptomycin (Sigma) per ml. Rabbit polyclonal antibodies raised against synthetic HXB_2 IN peptides (amino acids 23 to 34 and 142 to 153) were from Duane Grandgenett at St. Louis University through the AIDS Research and Reference Reagent Program. The sera from HIV-1-infected patients were purchased from The Scripps Research Institute.

Recombinant expression plasmids. The genes encoding the fusion proteins IN/E2C and E2C/IN were each cloned into a pLR2P expression plasmid, which was provided by Beatrice Hahn of the University of Alabama, Birmingham. The pLR2P construct encodes the HIV-1 accessory protein Vpr and contains an HIV-1 protease cleavage (PC) site following Vpr (20). To construct the R-PC-IN/E2C clone, an NsiI-SalI fragment of pIN1-288/E2C, which is derived from the pT7-7 expression vector and contains the coding sequence of full-length HIV-1 IN fused with E2C (57), was inserted in-frame into pLR2P previously digested with NsiI and XhoI. For the R-PC-E2C/IN clone, an NdeI (blunted)-SalI fragment of pE2C/IN1-288, which contains the coding sequence of E2C/IN fusion protein (57), was ligated with pLR2P previously cut with XhoI and partially digested with ScaI. Several constructs in which the E2C was fused to different truncated variants of HIV-1 IN were also prepared. To construct R-PC-IN1-234/E2C, an NsiI-SalI fragment of pIN1-234/E2C (57), which contains the coding sequence of the C terminus-truncated IN fused to E2C, was ligated with the pLR2P vector previously digested with NsiI and XhoI. Similarly, pR-PC-E2C/IN11-288 was prepared by inserting an NsiI (blunted)-SalI fragment of pE2C/IN11-288 (57), which contains the coding sequence of E2C fused to N terminus-truncated IN, into the pLR2P vector previously cut with XhoI and partially digested with ScaI. For pR-PC-IN50-288/E2C, a KpnI-SalI fragment of pIN1-288/E2C, which contains the coding sequence of E2C, was ligated with pR-PC-IN50-288/LA (27) previously cut with XhoI and partially digested with KpnI.

The HIV-1 plasmids pHXB-Hygro and pHXB-IN64-Hygro, which encode the wild-type (WT) and D64V mutated forms of IN, respectively, were obtained from Andrew D. Leavitt at the University of California, San Francisco. Both plasmids encode the HXB_2 strain of HIV-1 with the hygromycin resistance gene in place of the envelope (27). The expression plasmid pMD.G for the G glycoprotein of vesicular stomatitis virus (VSV-G) was provided by Didier Trono at the University of Geneva (45).

Virus preparation. All viral stocks were prepared by standard calcium phosphate transfection of monolayers of 293T cells with 20 μg of DNA in 75-cm^2 flasks (1). For virus stocks used in the hygromycin resistance assay, the hygromycin resistance construct with or without the IN mutation (pHXB-Hygro or pHXB-IN64-Hygro) was cotransfected with the pMD.G construct at a 3:1 ratio. The viruses provided with the integrase-E2C fusion proteins in *trans* were generated by triple transfection of pHIV-IN64-Hygro, pMD.G, and the appropriate pLR2P expression construct at a ratio of 5:1:14. Culture supernatants were collected 48 h after transfection. Virions were treated with 10 to 20 U of RNase-free DNase I (Amersham Pharmacia) per ml of viral stock at room temperature for 1 h and then pelleted by ultracentrifugation at 125,000 × *g* for 2 h at 4°C through a cushion of 20% sucrose. Viral pellets were resuspended in culture medium without serum and stored at −80°C until use. The virus titer was estimated by an enzyme-linked immunosorbent assay (Coulter Inc.) against the HIV-1 capsid (CA; p24) antigen.

Hygromycin resistance infectivity assay. The infectivity of various virus stocks was assessed using a hygromycin resistance assay as described previously (27). Five, 20, or 100 nanograms p24 equivalent of viruses collected from double (pHXB-Hygro or pHXB-IN64-Hygro and pMD.G) or triple (pHXB-IN64-Hygro, pMD.G, and a pLR2P expression construct) transfections were used to infect 1×10^6 HeLa cells in 100-mm plates. The medium containing the virus was removed 3 h postinfection and replaced with fresh DMEM. The cells were maintained in nonselective media for an additional 48 h. The DMEM was then exchanged for a medium containing 200 μg/ml hygromycin B (Sigma). Selection was continued for 21 days, and the colonies were then stained with 0.2% crystal violet in 10% phosphate-buffered formalin (pH 7.0) and counted.

Quantitative, two-step real-time PCR assay for proviral DNA. HIV-1 DNA integrated into the chromosome of infected cells was quantified using a fluorescence-monitored two-step PCR assay modified from those described previously (7, 47). All the primers and probes were synthesized by Operon Technologies Inc. and listed in Table 1.

(i) Cellular proviral DNA standard. To determine accurately the copy number of integrated proviruses in a cell sample, a DNA standard representative of a natural infection with a full distribution of distances between the provirus long terminal repeat (LTR) and the nearest *Alu* element is necessary. Cellular DNA containing proviruses integrated at many locations was generated by infecting HeLa cells with the HXB-IN64-Hygro virus containing Vpr-IN and pseudotyped with VSV-G at a multiplicity of infection of 10. The infected cells were grown in the presence of 200 μg/ml hygromycin B for four weeks to select cells containing proviruses and to allow the loss of unintegrated DNA. Total genomic DNA was isolated using the DNeasy tissue kit (QIAGEN). The copy number of proviral DNA per cell was determined by quantitative real-time PCR using 400 nM each of the primers MH531 (which anneals to nucleotide positions 556 to 576) and MH532 (which anneals to nucleotide positions 699 to 680) specific for late

TABLE 1. DNA sequences of primers and probes used in construction of fusion proteins and real-time PCR assays

Primer or probe	Sequence[a]
λT	5′-ATGCCACGTAAGCGAAACT
Alu1	5′-TCCCAGCTACTGGGGAGGCTGAGG
BGF	5′-CAACCTCAAACAGACACCATG
BGR	5′-TCCACGTTCACCTTGCCC
BGX-P	5′-FAM-CTCCTGAGGAGAAGTCTGCCGTTACTGCC-TAMRA
EcF1	5′-GAGGGCTGCTTGAGGAAGTATAAG
EcF2	5′-GGCCCTTTACTGCGCCGCGC
EcR1	5′-AGCCCCCAGCGGCACAAGGCCAGC
EcR2	5′-GGTCCCGCCGCTGCTCCGT
GKR	5′-GCTTGGTACCGACGCTCTCGCACCCA (KpnI)
LBF	5′-CGCGGATCCTGGAAGGGCTAATTTGGTC (BamHI)
LER	5′-GGTGATATCTGCTAGAGATTTTCCACAC (EcoRV)
LM652	5′-ATGCCACGTAAGCGAAACTCCCTGTTCGGGCGCCACTGCTAGAG
LM667	5′-ATGCCACGTAAGCGAAACTCTGGCTAACTAGGGAACCCACTG
LR	5′-TCCACACTGACTAAAAGGGTCTGA
LRT-P	5′-FAM-CAGTGGCGCCCGAACAGGGA-TAMRA
MH531	5′-TGTGTGCCCGTCTGTTGTGT
MH532	5′-GAGTCCTGCGTCGAGAGAGC
MH535	5′-AACTAGGGAACCCACTGCTTAAG
NXF	5′-TGGGAGCAGTATCTCGAGACCTAG (XhoI)
ZXF-P	5′-FAM-TGTGACTCTGGTAACTAGAGATCCCTCAGACCC-TAMRA

[a] Restriction sites are underlined, and the identity of each restriction enzyme is listed within the parentheses. FAM, 6-carboxyfluorescein; TAMRA, 6-carboxytetramethylrhodamine.

reverse transcriptase (RT) products (Table 1) as described previously (11). The fluorescent TaqMan probe LRT-P for the amplified U5-*gag*-containing product was modified at the 5′ and 3′ ends with 6-carboxyfluorescein (FAM) and 6-carboxytetramethylrhodamine (TAMRA), respectively. The probe anneals to nucleotide positions 633 to 652 of HIV-1 and was used at a concentration of 200 nM. PCR was performed using an ABI Prism 7900 sequence detection system and TaqMan Universal PCR master mix (PE-Applied Biosystems). The amplification condition included a hot start of 50°C for 2 min and 95°C for 10 min, followed by 40 cycles of denaturation at 95°C for 15 s and extension at 60°C for 1 min. A quantitative dilution series of pCR-5L-*e2c* (see below) was used as a copy number standard. The copy number of proviral DNA per cell was calculated, and the genomic DNA isolated from this standard cell line was referred as "*Alu*-PCR standard."

As an internal standard for normalizing the amount of cellular DNA, the level of human β-globin DNA was also measured using the TaqMan PCR reagent kit (PE-Applied Biosystems) following the manufacturer's instruction. The reaction was carried out using 400 nM of forward primer BGF, 400 nM of reverse primer BGR, and 200 nM of BGX-P as the fluorescence probe. The standard curve for the human β-globin sequence was generated using genomic DNA isolated from uninfected HeLa cells.

(ii) HIV/*e2c* DNA copy number standard. Given the impracticality of generating a DNA standard representative of HIV-1 DNA integrated randomly at a range of distances from the *e2c* site, we constructed two plasmids, pCR-5L-*e2c* and pCR-3L-*e2c*, as copy number standards for quantifying proviruses integrated downstream and upstream, respectively, of the *e2c* site. A 400-bp DNA fragment flanking the E2C-binding site, located within the 5′ untranslated region of the *erbB-2* gene in human chromosome 17 (3), was amplified with primers EcF1 and EcR1 using cellular DNA isolated from HeLa cells as the template. The PCR product was cloned into the pCR-Blunt II-Topo vector (Invitrogen), resulting in pCR-*e2c*. A DNA fragment containing the 5′ LTR and part of the *gag* gene of HIV-1 was obtained by PCR amplification using pHXB-Hygro as a template and LBF (which anneals to nucleotide positions 1 to 19) and GKR (which anneals to nucleotide positions 816 to 791) as primers. The PCR product was digested with BamHI and KpnI, purified by agarose gel electrophoresis, and then inserted into pCR-*e2c* previously digested with BamHI and KpnI to generate pCR-5L-*e2c* (see Fig. 4A). For pCR-3L-*e2c* (see Fig. 4A), a fragment containing the 3′ LTR and part of the *nef* gene of HIV-1 was amplified by PCR using pHXB-Hygro as a template and NXF (which anneals to nucleotide positions 8,875 to 8,898) and LER (which anneals to nucleotide positions 9,709 to 9,691) as primers. The PCR product was digested with XhoI and EcoRV, purified by agarose gel electrophoresis, and then ligated with pCR-*e2c* previously digested with XhoI and EcoRV. The sequences of all clones were verified by restriction analysis and Sanger DNA sequencing.

(iii) Quantitation of total integrated proviral DNA in the human genome by *Alu*-PCR. HeLa cells were infected with viruses containing WT IN or various IN-E2C fusion proteins in the presence of 20 μg/ml DEAE-dextran at a multiplicity of infection of 1. Genomic DNA was extracted with the DNeasy tissue kit (QIAGEN) three days after infection and diluted in Tris-EDTA buffer (pH 7.5). To quantify the total integration events in the human genome, an *Alu*-LTR-based real-time nested-PCR procedure was performed (*Alu*-PCR) (see Fig. 2). In the first round of PCR, proviral DNA sequences were amplified with an outward-facing *Alu* primer, Alu1, that anneals within conserved regions of *Alu* repeat elements and an HIV-1 LTR-specific primer, LM667 (which anneals to nucleotide positions 494 to 516). The 5′ end of the LTR-specific primer contains a phage lambda-specific heel sequence. *Alu*-LTR sequences were amplified in duplicates from 100 ng human genomic DNA in a 20-μl reaction mixture containing 10 mM Tris-HCl (pH 8.3), 1.5 mM $MgCl_2$, 1 mM deoxyribonucleoside triphosphates, 50 mM KCl, 100 nM Alu1 primer, 300 nM LM667 primer, and 0.5 U of Platinum *Taq* DNA polymerase (Invitrogen Life Technologies). The PCR cycle condition was as follows: a denaturation step of 5 min at 95°C and then 20 cycles of amplification (94°C for 15 s, 55°C for 30 s, and 70°C for 2 min). In the second round of PCR using the lambda-specific primer λT and an internal LTR primer, LR (which anneals to nucleotide positions 622 to 599), only proviral-containing products from the first-round PCR should be amplified. One-tenth of the first-round PCR product was used in a 20-μl mixture containing TaqMan Universal PCR master buffer, 400 nM λT primer, and 400 nM LR primer. The fluorescent TaqMan probe ZXF-P, which anneals to nucleotide positions 574 to 606 of HIV-1, was used at a concentration of 200 nM for the amplified R-U5-containing product. The reaction was performed on an ABI Prism 7900 sequence detection system with a thermal program as follows: a hot start of 50°C for 2 min, and 95°C for 10 min, followed by 40 cycles of denaturation at 95°C for 15 s and extension at 60°C for 1 min. The copy number of total integrated HIV-1 DNA in the cell sample was determined in reference to a standard curve generated by a concomitant *Alu*-LTR-based real-time nested PCR using 100 ng of *Alu*-PCR standard serially diluted with genomic DNA isolated from uninfected cells. A kinetic PCR assay for human β-globin DNA was also carried out to determine the cell number of each sample, as described earlier in "Cellular proviral DNA standard."

To control linear amplification arising from the LM667 primer, the nested-PCR procedure omitting the Alu1 primer in the first-round PCR was performed for each virus-infected sample. The copy number of integrated proviruses was then adjusted by subtracting the copy number quantified in the absence of the Alu1 primer (ranging from <0.1 to 0.7 copies/1,000 cells) from the copy number measured in the presence of the Alu1 primer. To test the specificity of the first-round PCR, the amplification was carried out with genomic DNA extracted from uninfected HeLa cells as the template and Alu1 and the HIV-1 LTR-specific LM667 as the primers. To test the specificity of the second-round PCR, the first-round PCR was carried out with *Alu*-PCR standard as the template and

Alu1 and LM667 without the phage lambda sequences as the primers. In both specificity controls, one-tenth of the first-round PCR product was used in the second round of PCR with the lambda-specific primer λT and an internal LTR primer, LR, under the identical condition as described earlier. The signal of both reactions was <0.1 copy/1,000 cells.

(iv) Quantitation of proviral DNA integrated near the *e2c* site by *e2c*-PCR. To quantify integration events near the E2C-binding site in the human genome, a real-time nested-PCR procedure (*e2c*-PCR) similar to the *Alu*-PCR described above was carried out (see Fig. 2). The Alu1 primer of the first-round PCR was replaced with *e2c*-specific primers EcF2 and EcR2, which anneals to sequences 67 bp upstream and 11 bp downstream, respectively, of the *e2c* site on human chromosome 17. The primers were chosen using the MacVector software program (Accelrys Inc.) and tested for low background signals in the quantitative PCR using uninfected human genomic DNA as the template (data not shown). To measure the number of proviruses integrated in both orientations upstream of the E2C-binding site, two primer sets were used in two separate first-round PCRs. One primer set included LM652 and EcR2 as the forward and reverse primers, respectively. The other primer set included LM667 and EcR2 as the forward and reverse primers, respectively. Both LM652 and LM667 contain a phage lambda-specific heel sequence and anneal to U5 and R regions of the HIV-1 LTR, respectively. Similarly, the number of proviruses integrated in both orientations downstream of the E2C-binding site was measured by two separate first-round PCRs. The forward primer in both reactions was EcF2 and the reverse primer was LM652 or LM667. The PCR condition was identical to that described earlier for *Alu*-PCR. In the second-round real-time PCR, one-tenth of the first-round PCR product was used as the template. In reactions where the primer LM652 was used in the first-round PCR, the forward and reverse primers for the second-round PCR were 400 nM λT and 400 nM MH535, respectively. MH535 anneals to nucleotide positions 500 to 522 of HIV-1 U5. In reactions where the primer LM667 was used in the first-round PCR, the forward and reverse primers for the second-round PCR were 400 nM λT and 400 nM LR, respectively. In both second-round reactions, 200 nM ZXF-P was used as the fluorescent TaqMan probe. We estimated that under the described PCR condition, each primer set could amplify DNA fragments that are on average 3 kbp in length. Therefore, the *e2c*-PCR assay is capable of measuring proviruses integrated within a 6-kbp region surrounding the E2C-binding site.

The copy number of integrated DNA was determined in reference to a standard curve generated by a concomitant two-step PCR amplification of a serial dilution of the *e2c*-PCR DNA standard pCR-5L-*e2c* for quantifying proviruses downstream of the *e2c* site, and pCR-3L-*e2c* for quantifying proviruses upstream of the *e2c* site. As in *Alu*-PCR, linear amplification arising from the LM667 and LM652 primers was determined for each virus-infected sample in the absence of the *e2c*-specific primers (ranging from <0.1 to 0.3 copies/1,000 cells) and subtracted from the copy number quantified in the presence of the *e2c*-specific primers. The quantitation of proviral DNA integrated near the *e2c* site of each sample was done concomitantly with the real-time PCR assay described earlier for human β-globin DNA.

RESULTS

Preparation of HIV-1 containing IN-E2C fusion proteins. To evaluate whether or not fusion proteins consisting of HIV-1 IN and designed polydactyl zinc finger protein E2C can mediate site-directed integration in vivo, we used the in *trans* method (26, 27, 63, 64) that has successfully incorporated various IN/LexA fusion proteins into infectious virions. In this method, the desired protein was linked to Vpr, an HIV-1 accessory protein that is packaged into viruses by interacting with the p6 protein of Gag (34, 42). The same strategy was used in the present study to incorporate WT IN and various HIV-1 IN-E2C fusion proteins into HIV-1 virions. After packaging, Vpr is removed from the IN-E2C fusion proteins by HIV-1 protease (20, 27). Viral stocks were prepared by transfecting 293T cells with three different plasmid DNAs (see Materials and Methods). The fusion protein plasmid expressed the various Vpr-IN-E2C fusion proteins, and the envelope plasmid expressed the VSV-G envelope protein for pseudotyping. The HIV-1 plasmid (HXB-IND64-Hygro) contains an Asp-to-Val mutation in residue 64 of IN, which produces a specific defect in catalyzing integration without affecting other critical stages of the virus life cycle (19, 36, 60). Since the IN encoded in the viral genome is catalytically inactive, the mutated clone allowed us to determine the activity of the various IN-E2C fusion proteins provided in *trans* and their ability to mediate site-directed integration.

A total of five fusion proteins containing E2C and full-length or truncation derivatives of HIV-1 IN were incorporated and tested. All five fusion proteins can bias integration of retroviral DNA near the E2C-binding site in vitro (57). IN/E2C and E2C/IN contained full-length HIV-1 IN at the N and C termini of E2C, respectively. E2C/IN11-288 and E2C/IN50-288 contained E2C with its C terminus fused to IN with 10- and 49-amino-acid deletions, respectively, at the N terminus. IN1-234/E2C contained a 54-amino-acid C terminus-truncated IN fused to the N terminus of E2C. As a positive control, the HXB-IND64-Hygro mutant virus was provided with WT IN in *trans*. The titer of the WT IN-containing virus was 270.3 ± 9.2 ng/ml. The titers of the fusion protein-containing viral clones, IN/E2C, E2C/IN, and E2C/IN11-288, as determined by measuring the level of p24, were 161.7 ± 6.6, 126.3 ± 3.5, and 266.3 ± 11.1 ng/ml, respectively. The titers of the two viruses, IN50-288/E2C and IN1-234/E2C, that contained fusion proteins with a 50-amino-acid truncation in the N and C termini of IN, respectively, were about 10-fold lower than that of the virus containing the WT IN. The basis for the titer reductions was not clear.

The presence of IN or various IN-E2C fusion proteins in virions was confirmed by Western blotting using anti-HIV-1 IN antibody. In all virus preparations, a predominant immunoreactive band corresponding presumably to WT, D64V IN, or both was detected (data not shown). For viruses containing each of the five IN-E2C fusion proteins, the anti-IN antibody detected proteins of the expected sizes, suggesting that the majority of each of the fusion proteins were processed appropriately by HIV-1 protease (data not shown). The results showed that in all cases the fusion proteins were incorporated into the virion and processed efficiently. Similar to previous observations, cleavage by either HIV-1 or cellular proteases did occur at additional sites in some of the incorporated fusion proteins (20, 27, 39, 63).

The assembly and packaging of the viral particles were also examined by immunoblotting with anti-HIV-1 human serum. For all viruses except those with E2C/IN11-288, incorporation of the fusion proteins into the virion did not appear to interfere with virus protein expression, packaging, and maturation (data not shown). Gag precursor proteins Pr55 and Pr41, along with the Gag processed product CA (p24) and the three processed products of the 160-kDa Gag-Pol precursor, IN (p32) and the p51 and p66 subunits of RT, were present (data not shown). For viruses with E2C/IN11-288, the levels of Pr55 and p66, and p32 were low and p51 was not detectable (data not shown). Whether the abnormality was due to a defect in packaging and maturation or selective protein degradation is unclear.

Integration of HIV-1 containing IN-E2C fusion proteins. To determine if viruses containing IN-E2C fusion proteins were infectious and capable of integrating their viral genome, a hygromycin resistance assay was performed. The assay uses an HIV-1 construct, HXB-IND64-Hygro, which has much of the

env gene replaced by a sequence encoding a selectable marker. The IN gene of the viral construct also contains a D64V mutation, which renders the virus unable to integrate. Integration activity of the D64V virus, therefore, depends solely on the functionality of the fusion protein provided in *trans*. If the fusion protein-containing virus is infectious and functional, synthesis and integration of viral DNA and subsequent expression of the hygromycin resistance gene in the provirus will allow clonal expansion of infected cells in the presence of the antibiotic.

Hygromycin-sensitive HeLa cells were infected with equal amounts of p24 equivalent of various viruses and then grown in media containing hygromycin B. For the negative control in which the virus contained the D64V IN mutation, no hygromycin-resistant colony was generated (Fig. 1a). For the positive control in which the D64V mutant virus was provided with WT IN in *trans* as a Vpr-IN fusion protein, an average of 12 resistant colonies were found per ng p24 equivalent of the virus (Fig. 1b). The numbers of resistant colonies produced by mutant viruses supplied with IN/E2C and E2C/IN were 24.2% and 11.0%, respectively (Fig. 1c and d), of that of the IN control. Hygromycin-resistant colonies were also generated by mutant viruses incorporated with truncated IN-E2C fusion proteins but at levels ranging from 0.8 to 1.4% of the IN control (Fig. 1e to g). The results confirmed that the in *trans*-produced viruses contained functional IN-E2C fusion proteins, which allowed the viral genome to be integrated and expressed in infected cells. These results are consistent with the previous work demonstrating that IN/LexA fusion proteins can be incorporated into infectious virions and restore integration to an IN-defective virus (26, 27).

Quantitative analysis of DNA integration events in the human genome by real-time PCR. To determine the ability of the IN-E2C fusion proteins to catalyze site-directed integration in culture cells, D64V IN mutant viruses provided with WT IN, IN/E2C, or E2C/IN fusion proteins in *trans* were used to infect cells and the levels of specific and total (specific plus nonspecific) integration events were quantified. Because of their low levels of infectivity (Fig. 1), mutant viruses containing the various truncated IN-E2C fusion proteins were not tested.

The number of proviruses integrated within the total genome and the number of proviruses at a specific loci were quantified using a real-time nested-PCR assay modified from those described previously (7, 47). Since integration of retroviral DNA is nonspecific and a considerable amount of unintegrated linear and circular DNA is present many days postinfection (7, 12), the total number of integrated proviruses within the genome of an infected cell was determined by quantifying the number of proviruses that integrated near *Alu* repeats (*Alu*-PCR) (Fig. 2). The *Alu* element was chosen as a PCR "anchor" for human genomic DNA because this element is the most numerous of repetitive DNA elements, with over one million copies randomly distributed and oriented in each diploid cell (35). The copy number of *Alu*-integrated proviral DNA was determined in reference to a standard curve generated concomitantly by the two-stage PCR amplification of a serial dilution of the *Alu*-PCR standard DNA (Fig. 3). Using the late RT primer set for quantifying HIV-1 DNA and the β-globin set for determining the cell number, we calculated

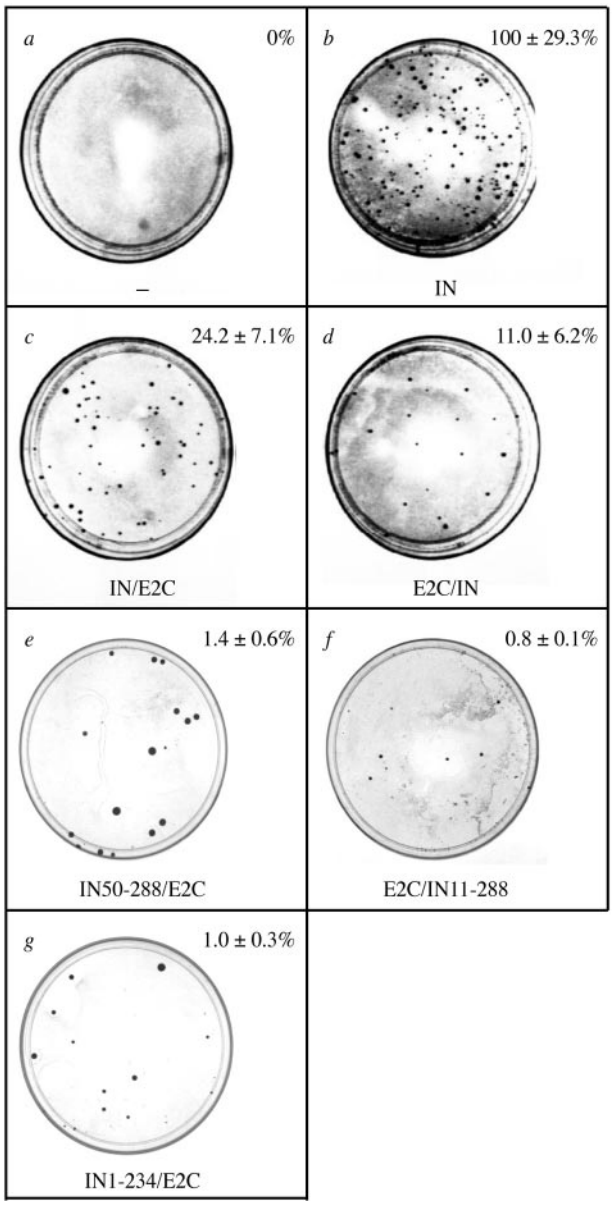

FIG. 1. Integration efficiency of viruses containing various IN-E2C fusion proteins. One million HeLa cells were infected for 4 h with equal amounts (ranging from 10 to 50 ng) of p24 equivalent of the indicated virus (panels a through g). Representative plates of the hygromycin resistance assay were shown. Dark spots on each plate are colonies that grew after selection with 200 μg/ml of hygromycin B for three weeks, beginning two days postinfection. The colonies are a result of provirus formation and stable expression of the hygromycin resistance gene. The number of resistant colonies per ng p24 for each virus was determined and expressed as a percentage of the colonies produced by the virus containing the WT IN in *trans* (panel b). Each determination at a particular p24 amount was performed in duplicate, and values on the upper-right corner of each panel are the mean ± standard error of the mean of three independent experiments. For the HXB-IN64 virus supplied with WT IN in *trans* (panel b), the number of resistant colonies per ng p24 was 11.6 ± 3.4.

that the *Alu*-PCR standard contained a mean of 1.8 proviruses per cell (data not shown).

To standardize the *Alu*-PCR for measuring the total integration events, a serially diluted DNA from the *Alu*-PCR stan-

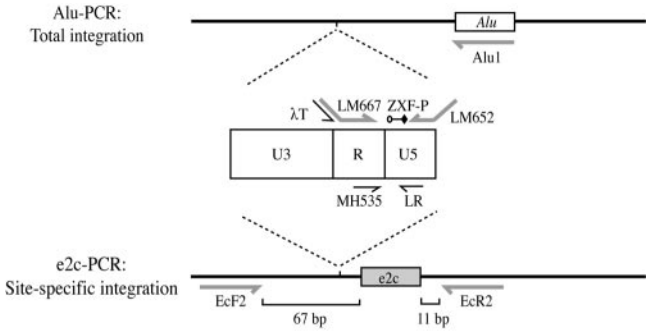

FIG. 2. A schematic diagram showing the real-time PCR assay for quantifying proviral DNA integrated in the entire genome (*Alu*-PCR) or near the E2C-binding site on chromosome 17 (*e2c*-PCR). HIV-1 DNA contains two LTRs, each located at the 5′ and 3′ ends of the viral genome. Integration of HIV-1 DNA can be upstream or downstream and in either orientations relative to a specified locus, *Alu* (open box) or *e2*c (shaded box). For simplicity, only one HIV-1 LTR and its subregions, U3, R, and U5, are shown, and the diagram depicts only the scenario in which the proviral DNA is integrated upstream of and in the same orientation as an *Alu* element or the *e2*c site. The quantitative assay consists of two rounds of PCR. Thick gray arrows represent the locations and orientations of the first-round PCR primers, whereas thin black arrows represent the locations and orientations of the second-round PCR primers. ZXF-P represents the location of the fluorescent probe. Primers LM652 and LM667 contain phage λ-specific sequences at their 5′ ends, and primer λT anneals specifically to phage λ sequences.

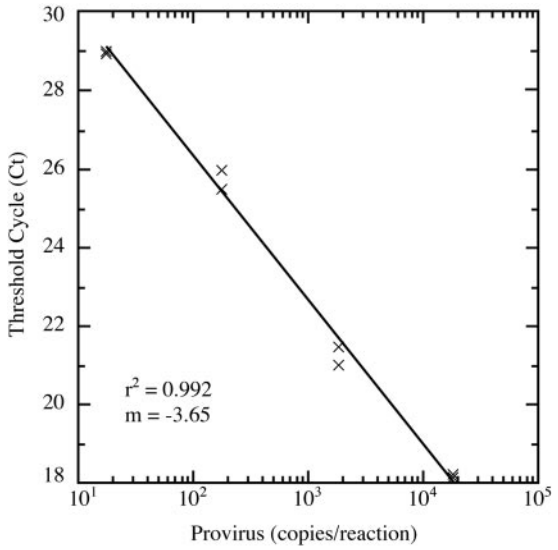

FIG. 3. A standard curve generated by *Alu*-PCR for measuring the total integration events in the human genome. *Alu*-PCR standard DNA, isolated from HeLa cells infected with HXB-IN64-Hygro/Vpr-IN, was serially diluted to contain 18 to 1.8×10^4 copies of HIV-1 proviruses per reaction. The proviral copy number of the *Alu*-PCR standard was previously determined using the pCR-5L-*e2c* as the standard and the late RT primers set (MH531 and MH532). Fluorescence curves of the *Alu*-PCR standard and infected DNA samples were generated after two-step amplification (data not shown), and the PCR cycles at which the amplification signal entered the exponential range were determined (threshold cycle [Ct]). The reactions were performed in duplicate.

dard was subjected to an initial nonkinetic *Alu*-LTR PCR and followed by a kinetic real-time PCR as described in Materials and Methods. The standard and samples were run concomitantly, and the standard curve generated by the real-time, two-stage PCR (Fig. 3) was used to calculate the total number of integrated proviruses in the infected cell samples. The linear regression obtained from amplifying serial dilutions of the *Alu*-PCR standard was linear over a 3-log range from 18 to 1.8×10^4 provirus copies per reaction, and the *Alu*-PCR assay could detect approximately 0.1 proviruses within 1,000 cell equivalents (Fig. 3).

The same fluorescence-monitored nested PCR described earlier can also be modified to quantify the number of proviruses integrated near the *e2c* site (*e2c*-PCR) (Fig. 2). Because the location of the *e2c* site is unique in the human genome, primers that anneal to sequences flanking the *e2c* site, rather than the *Alu* element, were used in the first-round PCR. Two separate sets of PCRs were carried out to measure integration events occurred downstream and upstream of the *e2c* site (Fig. 4). Graded doses of plasmids pCR-5L-*e2c* and pCR-3L-*e2c* (Fig. 4A) were used to obtain the standard curves for quantifying proviruses integrated downstream (Fig. 4B) and upstream (Fig. 4C), respectively, of the *e2*c site by the two-stage real-time PCR. The amplification signals for the pCR-5L-*e2c* DNA standard were linear over a 7-log range from 20 to 2×10^8 provirus copies per reaction, whereas the pCR-3L-*e2c* DNA standard amplification signals were linear over a 6-log range from 200 to 2×10^8 provirus copies per reaction.

As an internal standard for data comparison among different DNA samples, the human β-globin gene from each DNA sample was amplified under the identical condition described in Materials and Methods. The standard curve for the β-globin gene was linear over a 6-log range from 200 to 2×10^8 copies per reaction (data not shown).

HIV-1 containing IN-E2C fusion proteins integrates preferentially near the *e2c* site in human genome. To determine whether viruses containing IN-E2C fusion proteins are capable of directing integration into the predetermined *e2*c site in the human genome, cellular DNA was isolated from infected cells three days postinfection and analyzed using the two separate quantitative, real-time nested PCRs: *Alu*-PCR and *e2c*-PCR. The specificity of site-directed integration was quantified by comparing the number of proviruses integrated near the *e2*c site, as determined by the *e2c*-PCR, to that of the whole genome, as determined by the *Alu*-PCR.

For cells infected with the virus that incorporated the WT IN in *trans*, which served as a negative control for site-directed integration, we determined that on average, 212.7 copies of proviruses were present per 1,000 cells. Among these, an average of 0.31 copies of proviruses were integrated near the *e2*c site (Table 2). The percent total integrated proviruses found near the *e2c* site was therefore 0.15%. Consistent with their poorer integration activity in vitro (57) and infectivity in vivo (Fig. 1), cells infected with viruses provided with IN/E2C and E2C/IN fusion proteins had 59.8 and 8.1 copies, respectively, of total integrated proviruses per 1,000 cells (Table 2). However, the copy number of proviruses near the *e2*c site was proportionally higher for viruses containing IN-E2C fusion proteins than those with the WT IN. The IN/E2C-containing virus had 0.59 copies of proviruses per 1,000 cells, or 0.99% of the total, integrated near the *e2*c site; whereas the E2C/IN-containing

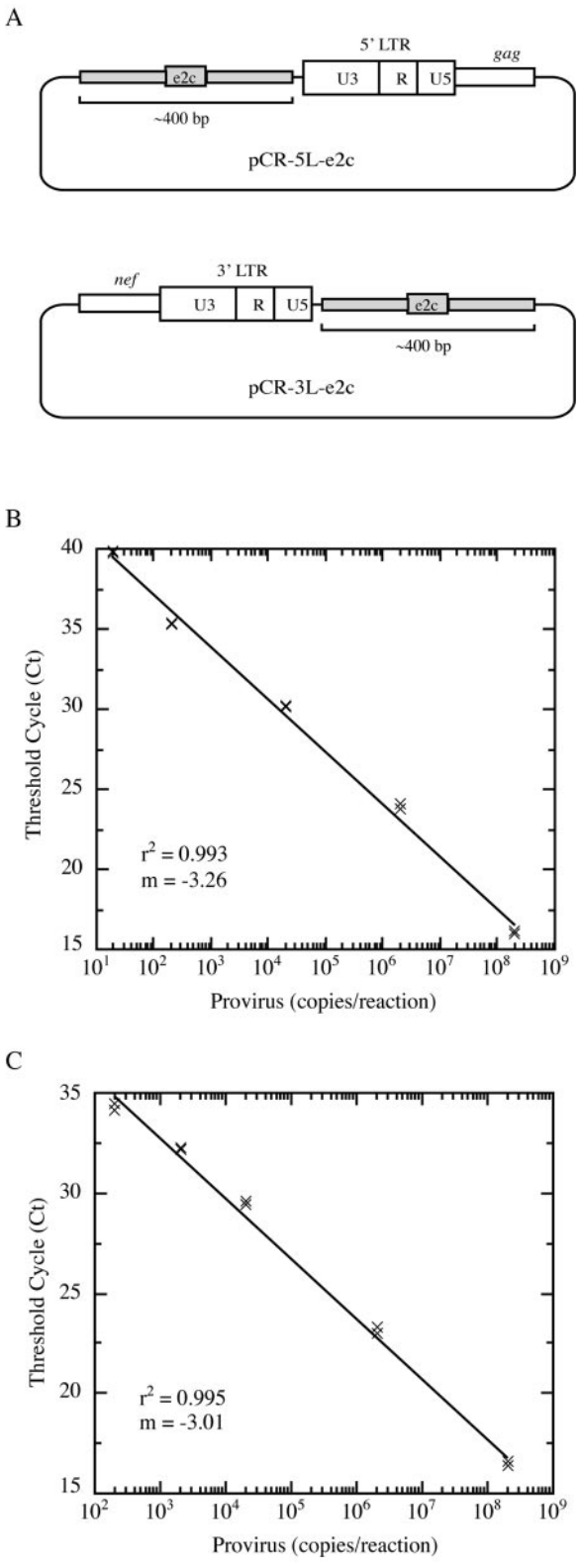

FIG. 4. DNA constructs and standard curves generated by *e2c*-PCR for measuring proviral DNA integrated upstream or downstream of the E2C-binding site. (A) DNA constructs. An HIV-1 DNA fragment (open box) containing the 5′ LTR and part of the *gag* sequence or the 3′ LTR and part of the *nef* sequence was obtained by PCR. The 5′ LTR-*gag* fragment was cloned into pCR-*e2c* downstream of the 400-bp *e2*c-containing fragment (shaded box), resulting in pCR-5L-*e2c*. The 3′ LTR-*nef* fragment was cloned upstream of the 400-bp *e2*c-containing fragment to form pCR-3L-*e2c*. (B) A standard curve generated by *e2c*-PCR for quantifying proviral DNA copies downstream of the E2C-binding site. Quantitative dilution series of pCR-5L-*e2c* DNA ranging from 20 to 2.0×10^8 copies per reaction were prepared, and fluorescence curves were generated by two-step amplification (data not shown). Logarithmic regression of the threshold cycles for the various provirus copies per reaction was plotted for quantifying HIV-1 proviral DNA. The reactions were performed in duplicate. (C) A standard curve generated by *e2c*-PCR for quantifying proviral DNA copies upstream of the E2C-binding site. The reaction was performed as described in panel B except that a serial dilution of pCR-3L-*e2c* DNA ranging from 200 to 2.0×10^8 copies per reaction was prepared as the DNA standard.

virus had 0.12 copies of proviruses per 1,000 cells, or 1.48% of the total, integrated near the *e2*c site. If we arbitrarily set the percent of *e2*c-specific integration of the WT IN-containing virus as 1, the relative specificities of *e2*c-directed integration of viruses containing IN/E2C and E2C/IN fusion proteins were 6.6- and 9.9-fold, respectively, higher than that of the WT IN. The result indicated that viruses provided with IN-E2C fusion proteins are capable of directing integration of viral DNA into a predetermined region in the human genome.

DISCUSSION

Once integrated, the DNA genome of retroviruses is stably transmitted as an integral element of the host cell genome. The ability to permanently and precisely insert a gene of interest into the chromosome of a target cell makes retroviruses a desirable vector for genetic engineering and therapy in mammalian cells (21, 58, 59). Integration of retroviral DNA is, however, nonspecific and can occur at many sites throughout the chromosomes. The nonspecific nature of integration raises a safety concern that may limit the usefulness of retrovirus-based vectors in gene delivery. Nonspecific integration into different chromosomal locations may also lead to variations in the expression level of a transgene (28). Therefore, the use and safety of retrovirus-based vectors can be greatly improved if the integration event is directed into a specific region. In vitro, selection of target sites can be favorably manipulated by fusing retroviral IN with a sequence-specific DNA-binding protein, such as LexA, λ repressor, Zif268, and the designed poly-zinc finger protein E2C (9, 10, 22, 31, 57). Whether or not the IN fusion strategy can function in vivo has not been determined. By quantifying integration events in cultured human cells by using fluorescence-monitored real-time PCR, we showed that viruses supplied with the IN-E2C fusion protein in *trans* can bias the delivery of a gene of interest into a predetermined DNA region.

E2C is a designed six-zinc finger protein that binds with high affinity to a contiguous 18-bp sequence (16), which is uniquely located within the 5′ untranslated region of the *erbB-2* gene in human chromosome 17 (4). E2C also has a modular organization and the fingers may be mixed and matched to provide proteins with new DNA-binding specificities (4, 14, 41, 51, 55). Although site-directed integration has been reported previously in vitro with fusion proteins consisting of IN and various sequence-specific binding proteins (9, 10, 22, 31), the use of E2C and other polydactyl zinc-finger protein as the target-

TABLE 2. Quantitative analysis of site-directed integration in the human genome

Virus construct	No. of proviruses/10^3 cells[a]		Ratio[b]	Relative specificity
	Total (*Alu*-PCR)	Site specific (*e2c*-PCR)		
HXB-IN64:IN	212.7 (212.6, 212.8)	0.31 (0.33, 0.29)	0.0015	1.0
HXB-IN64:IN/E2C	59.8 (58.5, 61.0)	0.59 (0.74, 0.44)	0.0099	6.6
HXB-IN64:E2C/IN	8.1 (7.6, 8.5)	0.12 (0.11, 0.13)	0.0148	9.9

[a] Values are average of two independent experiments. The numbers within the parentheses are values of each individual experiment. All reactions in each experiment were performed in duplicate.
[b] Ratios are as follows: proviral copy number determined by *e2c*-PCR/proviral copy number determined by *Alu*-PCR.

specifying component offers important advantages in terms of specificity and versatility. Also, because the location of the E2C-recognition site is unique, the number of proviruses specifically targeted in the vicinity of the binding site can be assayed using a fluorescence-monitored nested-PCR assay (7, 47). Our analysis showed that, in comparison with the WT, viruses incorporated with IN/E2C and E2C/IN had a 7- and 10-fold increase, respectively, in the copy number of proviral DNA integrated near the E2C-binding site. Although the basis is presently unknown, the higher integration preference displayed by viruses containing E2C/IN is consistent with the higher integration specificity of the purified E2C/IN than that of IN/E2C in vitro (57). In addition to the quantitative PCR assay, we are currently cloning and sequencing individual integration sites from the various infected samples to analyze the genome-wide distribution of proviruses and to further confirm the integration bias of the IN-E2C fusion proteins.

The various IN-E2C fusion proteins were incorporated into viral particles using the in *trans* method (27, 64), which involves linking the protein of interest to Vpr, an HIV-1 accessory protein that is packaged into viruses by interacting with the p6 protein of Gag (34). All resulting viruses were infectious, contained the intended IN-E2C fusion proteins, and stably expressed a selectable marker gene. However, based on the hygromycin resistance assay, the ability of the IN-E2C fusion protein to restore integration was at best 24% of that of the WT IN. It is possible that the IN-E2C fusion proteins have a poorer incorporation efficiency, higher level of proteolysis, or lower integration activity than that of the WT IN. A simpler alternative approach to the in *trans* method in packaging the fusion protein into virions is to place the fusion gene directly into the viral genome. Previous attempts to produce retroviruses containing IN-fusion proteins by encoding them in the viral genome were unsuccessful due to loss of viral infectivity after transfection (10) or loss of fusion protein expression during viral replication owing to reversion (31). Each of these attempts used a fusion protein with the sequence-specific DNA-binding protein tethered to the C terminus of IN. The addition of the DNA-binding protein to the C terminus of IN may disrupt the folding, packaging, or processing of the Gag-Pol polyprotein. Another possible difficulty in the case of HIV-1 lies in the fact that the 3′ end of the IN gene overlaps the coding sequence of *vif* and contains a splice acceptor site (49). Substituting the IN gene with that of the fusion protein may disrupt the critical elements of the viral genome. Since the fusion proteins containing E2C at either terminus of IN are effective (Fig. 1 and Table 2) (57), a viral construct in which the IN gene is substituted with the E2C/IN fusion gene will be tested for its ability to produce infectious virions and to express the E2C/IN fusion protein. The protease cleavage site at the 3′ end of RT will be maintained to avoid interference with the correct processing of RT and IN. With the E2C spliced to the 5′ end of IN, we reason that this construct will have no effect on *vif* and the splice acceptor site and may result in a fusion protein-expressing virus with a higher integration efficiency.

Integration of retroviral DNA is inherently a mutagenic event and can deregulate cellular gene expression by insertional mutagenesis. The importance of insertional mutagenesis is further highlighted by reports of leukemia in mice and infants after the administration of retrovirus-based vectors (24, 38). The IN-E2C fusion proteins are catalytically active and can be stably expressed and incorporated into infectious virions. The results demonstrate that fusion proteins consisting of IN and designed polydactyl zinc finger protein can direct integration of retroviral DNA into a predetermined DNA region in cultured cells, providing a feasible and versatile approach for delivering genes for genetic engineering and therapeutic purposes.

ACKNOWLEDGMENTS

We thank David Segal for helpful discussions, Jerome Zack for comments and critical reading of the manuscript, and the Core Virology Laboratory at the UCLA AIDS Institute for carrying out the enzyme-linked immunosorbent assay for p24.

This work was supported by National Institutes of Health grants CA68859 to S.A.C. and GM065059 to C.F.B. and NIAID postdoctoral training grant T32 AI060567 to T.A.W.

REFERENCES

1. **Ausubel, F. A., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl.** 1999. Current protocols in molecular biology. Wiley, New York, N.Y.
2. **Beerli, R. R., and C. F. Barbas III.** 2002. Engineering polydactyl zinc-finger transcription factors. Nat. Biotechnol. **20:**135–141.
3. **Beerli, R. R., B. Dreier, and C. F. Barbas III.** 2000. Positive and negative regulation of endogenous genes by designed transcription factors. Proc. Natl. Acad. Sci. USA **97:**1495–1500.
4. **Beerli, R. R., D. J. Segal, B. Dreier, and C. F. Barbas, III.** 1998. Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. Proc. Natl. Acad. Sci. USA **95:**14628–14633.
5. **Blancafort, P., D. J. Segal, and C. F. Barbas III.** 2004. Designing transcription factor architectures for drug discovery. Mol. Pharmacol. **66:**1361–1371.
6. **Brown, P. O.** 1997. Integration, p. 161–204. *In* J. M. Coffin, S. H. Hughes, and H. E. Varmus (ed.), Retroviruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
7. **Brussel, A., and P. Sonigo.** 2003. Analysis of early human immunodeficiency virus type 1 DNA synthesis by use of a new sensitive assay for quantifying integrated provirus. J. Virol. **77:**10119–10124.
8. **Bushman, F. D.** 2002. Integration site selection by lentiviruses: biology and possible control. Curr. Top. Microbiol. Immunol. **261:**165–177.
9. **Bushman, F. D.** 1994. Tethering human immunodeficiency virus 1 integrase to a DNA site directs integration to nearby sequences. Proc. Natl. Acad. Sci. USA **91:**9233–9237.

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. **Butler, S. L., M. S. Hansen, and F. D. Bushman.** 2001. A quantitative assay for HIV DNA integration in vivo. Nat. Med. **7:**631–634.
12. **Butler, S. L., E. P. Johnson, and F. D. Bushman.** 2002. Human immunodeficiency virus cDNA metabolism: notable stability of two-long terminal repeat circles. J. Virol. **76:**3739–3747.
13. **Choo, Y., A. Castellanos, B. Garcia-Hernandez, I. Sanchez-Garcia, and A. Klug.** 1997. Promoter-specific activation of gene expression directed by bacteriophage-selected zinc fingers. J. Mol. Biol. **273:**525–532.
14. **Choo, Y., and A. Klug.** 1994. Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage. Proc. Natl. Acad. Sci. USA **91:**11163–11167.
15. **Craigie, R.** 1992. Hotspots and warm spots: integration specificity of retroelements. Trends Genet. **8:**187–190.
16. **Desjarlais, J. R., and J. M. Berg.** 1994. Length-encoded multiplex binding site determination: application to zinc finger proteins. Proc. Natl. Acad. Sci. USA **91:**11099–11103.
17. **Desjarlais, J. R., and J. M. Berg.** 1993. Use of a zinc-finger consensus sequence framework and specificity rules to design specific DNA binding proteins. Proc. Natl. Acad. Sci. USA **90:**2256–2260.
18. **Dreier, B., R. R. Beerli, D. J. Segal, J. D. Flippin, and C. F. Barbas, III.** 2001. Development of zinc finger domains for recognition of the 5′-ANN-3′ family of DNA sequences and their use in the construction of artificial transcription factors. J. Biol. Chem. **276:**29466–29478.
19. **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.
20. **Fletcher, T. M., III, M. A. Soares, S. McPhearson, H. Hui, M. Wiskerchen, M. A. Muesing, G. M. Shaw, A. D. Leavitt, J. D. Boeke, and B. H. Hahn.** 1997. Complementation of integrase function in HIV-1 virions. EMBO J. **16:**5123–5138.
21. **Gaiano, N., and N. Hopkins.** 1996. Introducing genes into zebrafish. Biochim. Biophys. Acta **1288:**11–14.
22. **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.
23. **Guan, X., J. Stege, M. Kim, Z. Dahmani, N. Fan, P. Heifetz, C. F. Barbas III, and S. P. Briggs.** 2002. Heritable endogenous gene regulation in plants with designed polydactyl zinc finger transcription factors. Proc. Natl. Acad. Sci. USA **99:**13296–13301.
24. **Hacein-Bey-Abina, S., C. von Kalle, M. Schmidt, F. Le Deist, N. Wulffraat, E. McIntyre, I. Radford, J. L. Villeval, C. C. Fraser, M. Cavazzana-Calvo, and A. Fischer.** 2003. A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. N. Engl. J. Med. **348:**255–256.
25. **Holmes-Son, M. L., R. S. Appa, and S. A. Chow.** 2001. Molecular genetics and target site specificity of retroviral integration. Adv. Genet. **43:**33–69.
26. **Holmes-Son, M. L., and S. A. Chow.** 2002. Correct integration mediated by integrase-LexA fusion proteins incorporated into HIV-1. Mol. Ther. **5:**360–370.
27. **Holmes-Son, M. L., and S. A. Chow.** 2000. Integrase-LexA fusion proteins incorporated into human immunodeficiency virus type 1 that contains a catalytically inactive integrase gene are functional to mediate integration. J. Virol. **74:**11548–11556.
28. **Jordan, A., P. Defechereux, and E. Verdin.** 2001. The site of HIV-1 integration in the human genome determines basal transcriptional activity and response to Tat transactivation. EMBO J. **20:**1726–1738.
29. **Jordan, S. R., and C. O. Pabo.** 1988. Structure of the lambda complex at 2.5 Å resolution: details of the repressor-operator interactions. Science **242:**893–899.
30. **Kang, J. S., and J. S. Kim.** 2000. Zinc finger proteins as designer transcription factors. J. Biol. Chem. **275:**8742–8748.
31. **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.
32. **Katz, R. A., and A. M. Skalka.** 1994. The retroviral enzymes. Annu. Rev. Biochem. **63:**133–173.
33. **Kim, J. S., and C. O. Pabo.** 1997. Transcriptional repression by zinc finger peptides. Exploring the potential for applications in gene therapy. J. Biol. Chem. **272:**29795–29800.
34. **Kondo, E., F. Mammano, E. A. Cohen, and H. G. Gottlinger.** 1995. The p6 gag domain of human immunodeficiency virus type 1 is sufficient for the incorporation of Vpr into heterologous viral particles. J. Virol. **69:**2759–2764.
35. **Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, et al.** 2001. Initial sequencing and analysis of the human genome. Nature **409:**860–921.
36. **Leavitt, A. D., G. Robles, N. Alesandro, and H. E. Varmus.** 1996. Human immunodeficiency virus type 1 integrase mutants retain in vitro integrase activity yet fail to integrate viral DNA efficiently during infection. J. Virol. **70:**721–728.
37. **Lewis, L. K., G. R. Harlow, L. A. Gregg-Jolly, and D. W. Mount.** 1994. Identification of high affinity binding sites for LexA which define new DNA damage-inducible genes in *Escherichia coli*. J. Mol. Biol. **241:**507–523.
38. **Li, Z., J. Dullmann, B. Schiedlmeier, M. Schmidt, C. von Kalle, J. Meyer, M. Forster, C. Stocking, A. Wahlers, O. Frank, W. Ostertag, K. Kuhlcke, H. G. Eckert, B. Fehse, and C. Baum.** 2002. Murine leukemia induced by retroviral gene marking. Science **296:**497.
39. **Liu, H., X. Wu, H. Xiao, J. A. Conway, and J. C. Kappes.** 1997. Incorporation of functional human immunodeficiency virus type 1 integrase into virions independent of the Gag-Pol precursor protein. J. Virol. **71:**7704–7710.
40. **Liu, P. Q., E. J. Rebar, L. Zhang, Q. Liu, A. C. Jamieson, Y. Liang, H. Qi, P. X. Li, B. Chen, M. C. Mendel, X. Zhong, Y. L. Lee, S. P. Eisenberg, S. K. Spratt, C. C. Case, and A. P. Wolffe.** 2001. Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. Activation of vascular endothelial growth factor A. J. Biol. Chem. **276:**11323–11334.
41. **Liu, Q., D. J. Segal, J. B. Ghiara, and C. F. Barbas III.** 1997. Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. Proc. Natl. Acad. Sci. USA **94:**5525–5530.
42. **Lu, Y. L., R. P. Bennett, J. W. Wills, R. Gorelick, and L. Ratner.** 1995. A leucine triplet repeat sequence (LXX)4 in $p6^{gag}$ is important for Vpr incorporation into human immunodeficiency virus type 1 particles. J. Virol. **69:**6873–6879.
43. **Marchetti, A., F. Buttitta, S. Miyazaki, D. Gallahan, G. H. Smith, and R. Callahan.** 1995. Int-6, a highly conserved, widely expressed gene, is mutated by mouse mammary tumor virus in mammary preneoplasia. J. Virol. **69:**1932–1938.
44. **Mitchell, R. S., B. F. Beitzel, A. R. Schroder, P. Shinn, H. Chen, C. C. Berry, J. R. Ecker, and F. D. Bushman.** 2004. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. PLoS Biol. **2:**1127–1137.
45. **Naldini, L., U. Blomer, P. Gallay, D. Ory, R. Mulligan, F. H. Gage, I. M. Verma, and D. Trono.** 1996. In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. Science **272:**263–267.
46. **Nusse, R.** 1991. Insertional mutagenesis in mouse mammary tumorigenesis. Curr. Top. Microbiol. Immunol. **171:**43–65.
47. **O'Doherty, U., W. J. Swiggard, D. Jeyakumar, D. McGain, and M. H. Malim.** 2002. A sensitive, quantitative assay for human immunodeficiency virus type 1 integration. J. Virol. **76:**10942–10950.
48. **Pavletich, N. P., and C. O. Pabo.** 1991. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. Science **252:**809–817.
49. **Purcell, D. F., and M. A. Martin.** 1993. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. J. Virol. **67:**6365–6378.
50. **Raineri, I., and H. P. Senn.** 1992. HIV-1 promotor insertion revealed by selective detection of chimeric provirus-host gene transcripts. Nucleic Acids Res. **20:**6261–6266.
51. **Rebar, E. J., H. A. Greisman, and C. O. Pabo.** 1996. Phage display methods for selecting zinc finger proteins with novel DNA-binding specificities. Methods Enzymol. **267:**129–149.
52. **Rebar, E. J., Y. Huang, R. Hickey, A. K. Nath, D. Meoli, S. Nath, B. Chen, L. Xu, Y. Liang, A. C. Jamieson, L. Zhang, S. K. Spratt, C. C. Case, A. Wolffe, and F. J. Giordano.** 2002. Induction of angiogenesis in a mouse model using engineered transcription factors. Nat. Med. **8:**1427–1432.
53. **Sandmeyer, S.** 2003. Integration by design. Proc. Natl. Acad. Sci. USA **100:**5586–5588.
54. **Schroder, A. R., P. Shinn, H. Chen, C. Berry, J. R. Ecker, and F. Bushman.** 2002. HIV-1 integration in the human genome favors active genes and local hotspots. Cell **110:**521–529.
55. **Segal, D. J., R. R. Beerli, P. Blancafort, B. Dreier, K. Effertz, A. Huber, B. Koksch, C. V. Lund, L. Magnenat, D. Valente, and C. F. Barbas III.** 2003. Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. Biochemistry **42:**2137–2148.
56. **Shiramizu, B., B. G. Herndier, and M. S. McGrath.** 1994. Identification of a

common clonal human immunodeficiency virus integration site in human immunodeficiency virus-associated lymphomas. Cancer Res. **54:**2069–2072.
57. **Tan, W., K. Zhu, D. J. Segal, C. F. Barbas III, and S. A. Chow.** 2004. Fusion proteins consisting of human immunodeficiency virus type 1 integrase and the designed polydactyl zinc finger protein E2C direct integration of viral DNA into specific sites. J. Virol. **78:**1301–1313.
58. **Temin, H. M.** 1990. Safety considerations in somatic gene therapy of human disease with retrovirus vectors. Hum. Gene Ther. **1:**111–123.
59. **Verma, I. M., and N. Somia.** 1997. Gene therapy—promises, problems and prospects. Nature **389:**239–242.
60. **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.
61. **Withers-Ward, E. S., Y. Kitamura, J. P. Barnes, and J. M. Coffin.** 1994. Distribution of targets for avian retrovirus DNA integration in vivo. Genes Dev. **8:**1473–1487.
62. **Wu, X., Y. Li, B. Crise, and S. M. Burgess.** 2003. Transcription start regions in the human genome are favored targets for MLV integration. Science **300:**1749–1751.
63. **Wu, X., H. Liu, H. Xiao, J. A. Conway, E. Hunter, and J. C. Kappes.** 1997. Functional RT and IN incorporated into HIV-1 particles independently of the Gag/Pol precursor protein. EMBO J. **16:**5113–5122.
64. **Wu, X., H. Liu, H. Xiao, J. Kim, P. Seshaiah, G. Natsoulis, J. D. Boeke, B. H. Hahn, and J. C. Kappes.** 1995. Targeting foreign proteins to human immunodeficiency virus particles via fusion with Vpr and Vpx. J. Virol. **69:**3389–3398.