JOURNAL OF VIROLOGY, Jan. 1996, p. 37–46 Vol. 70, No. 1
0022-538X/96/$04.00+0

Directed Integration of Viral DNA Mediated by Fusion Proteins Consisting of Human Immunodeficiency Virus Type 1 Integrase and *Escherichia coli* LexA Protein

HÉLÈNE GOULAOUIC[1] AND SAMSON A. CHOW[1,2]*

Department of Molecular and Medical Pharmacology,[1] and Molecular Biology Institute,[2] UCLA School of Medicine, Los Angeles, California 90095

Received 31 July 1995/Accepted 22 September 1995

We tested whether the selection of target sites can be manipulated by fusing retroviral integrase with a sequence-specific DNA-binding protein. A hybrid protein that has the *Escherichia coli* LexA protein fused to the C terminus of the human immunodeficiency virus type 1 integrase was constructed. The fusion protein, IN1-288/LA, retained the catalytic activities in vitro of the wild-type human immunodeficiency virus type 1 integrase (WT IN). Using an in vitro integration assay that included multiple DNA fragments as the target DNA, we found that IN1-288/LA preferentially integrated viral DNA into the fragment containing a DNA sequence specifically bound by LexA protein. No bias was observed when the LexA-binding sequence was absent, when the fusion protein was replaced by WT IN, or when LexA protein was added in the reaction containing IN1-288/LA. A majority of the integration events mediated by IN1-288/LA occurred within 30 bp of DNA flanking the LexA-binding sequence. The specificity toward the LexA-binding sequence and the distribution and frequency of target site usage were unchanged when the integrase component of the fusion protein was replaced with a variant containing a truncation at the N or C terminus or both, suggesting that the domain involved in target site selection resides in the central core region of integrase. The integration bias observed with the integrase-LexA hybrid shows that one effective means of altering the selection of DNA sites for integration is by fusing integrase to a sequence-specific DNA-binding protein.

Integration of retroviral DNA into the host cell genome is an essential step during the life cycle of retroviruses (50). Two factors are required for the integration process: the viral protein integrase and sequences at each end of the linear viral DNA. Integrase catalyzes the removal of two nucleotides from the 3′ end of each viral strand and the subsequent joining of the processed 3′ viral ends to the 5′-phosphates on each strand of target DNA in a concerted cleavage-joining reaction (for reviews, see references 6, 18, and 54).

In vivo and in vitro studies show that integration of retroviral DNA can occur into many sites on target DNA (reference 10 and references therein). The process, however, is not entirely random; the frequency of use of specific sites varies considerably, with some sites being preferred at a frequency up to 100 times greater than random (38, 51, 58). The mechanism that determines target site specificity is not well understood, and several factors that can affect target site selection, including DNA and chromatin structure, DNA methylation, DNA sequences, and DNA-binding proteins, have been identified. Integration occurs preferentially into regions near DNase I-hypersensitive sites and transcriptionally active genes (38, 51) and into runs of CpG islands modified by 5-methylation of cytosine (25). One factor important for target site selection that has been well characterized is chromatin structure. Nucleosomal DNA in the chromatin is preferred to nucleosome-free DNA, and integration tends to cluster in the exposed face of the major groove within the nucleosome core (33, 36). The basis for preferred integration in nucleosomes may be related to DNA distortion, as DNA bending itself creates favored sites for integration (31, 34). Although sequence analysis of integration sites has revealed only weak consensus sequences (16, 20), comparisons of the integration patterns in a DNA sequence in vivo and as a naked DNA in vitro show that the DNA sequence is also an important determinant in target site selection (35, 36).

Another factor in target site selection is sequence- or structure-specific DNA-binding proteins. Certain DNA-binding proteins, such as the yeast transcriptional repressor α2 and the *lac* repressor of *Escherichia coli*, can prevent integration, presumably by steric hindrance (31, 36). Unlike histones and other proteins that stimulate integration by inducing DNA bends, certain DNA-binding proteins may promote integration by interacting with the integration machinery. The significance of such an interaction is illustrated by the position-specific integration of the yeast retrovirus-like element Ty3 (39), which is mediated by a specific interaction between integrase and RNA polymerase III transcription factors (24). Among retroviruses, a stimulation of integration produced by a specific interaction between human immunodeficiency virus type 1 (HIV-1) integrase and a putative transcription activator led to the suggestion that such an interaction may target the viral DNA to active genes (22).

In addition to altering preferred target sites by interacting with cellular proteins, integrase itself is a major factor in determining target site specificity. Integration reactions carried out with purified integrase or integration complexes isolated from virus-infected cells show similar patterns of target specificity. However, there are significant differences in the distribution and frequency of integration sites between integrases of HIV-1 and murine leukemia virus (36) and of HIV-1 and feline immunodeficiency virus (44). The C-terminal third of integrase, the least conserved region among retroviral integrases (21), possesses DNA-binding activity (15, 40, 53, 59). The DNA binding by the C terminus does not show any sequence specificity, which led to its proposed role as the domain for binding target DNA, and this binding may partly explain the ability of integrase to insert viral DNA at sites with weak consensus sequences.

* Corresponding author. Phone: (310) 825-9600. Fax: (310) 825-6267. Electronic mail address: schow@pharm.medsch.ucla.edu.

In the absence of information on the mechanistic basis for target site selection by integrase, one experimental approach of altering the target specificity of retroviral integration is to fuse the integrase with a sequence-specific DNA-binding protein. Directed integration has been demonstrated by tethering integrase to a target DNA site, accomplished by use of a hybrid protein composed of a full-length HIV-1 integrase at the C terminus and the DNA-binding domain of λ repressor at the N terminus (7). The hybrid protein mediates integration preferentially to target DNA containing λ operators. The integration sites are near the λ operator on the same face of the DNA helix, indicating that the hybrid protein binds to the operator and captures targets probably by looping out the intervening DNA (7).

In this study, we independently have taken a similar approach of using a fusion protein, consisting of HIV-1 integrase and a sequence-specific DNA-binding protein, the *E. coli* LexA protein, to test the feasibility of directing integration of viral DNA into specific sites recognized by LexA protein. We also tested whether the specificity of integration can be enhanced by manipulating the integrase component of the fusion protein. We found that integration mediated by the integrase-LexA fusion proteins was biased toward target DNA containing the LexA-binding sequence and that deletion of approximately 50 amino acids at the N or C terminus or both termini of integrase had little effect on target specificity.

MATERIALS AND METHODS

DNA constructs. The HIV-1 integrase and the *lexA* genes were obtained from plasmids pT7-7-IN (52) and pBTM117 (kindly provided by John Colicelli), respectively. The genes were amplified by PCR. Oligonucleotide primers used in PCR were from Operon Technologies, Inc. The primers for the N terminus of the full-length and the N-terminus-truncated (amino acid residues 1 to 49) integrases were 5′-GAAGGAGATATACATATGTTTTTAGATGGA-3′ and 5′-TAGACTCATATGCATGGACAAGTA-3′, respectively. The N-terminus primers contain an *Nde*I site (underlined). The primers for the C terminus of the full-length and the C-terminus-truncated (amino acid residues 235 to 288) integrases were 5′-GCTAGAGGTACCATCCTCATCCTGTCTACT-3′ and 5′-GCTAGAGGTACCAACTGGATCTCTGCTGTC-3′, respectively. The C-terminus primers contain a *Kpn*I site (underlined). The primer for the N terminus of the *lexA* gene was 5′-CAGTCAGGTACCAAAGCGTTAACGGCCAGG-3′ and contains a *Kpn*I site (underlined). The primers for the C terminus of full-length LexA and the DNA-binding domain (amino acids 1 to 87) of LexA were 5′-ATAGGATCC*TTA*CAGCCAGTCGCCGTTGCG-3′ and 5′-ATTGGATCC*TTA*TGGTTCACCGGCAGC-3′, respectively. The C-terminus primers for the *lexA* gene contain a *Bam*HI site (underlined) and a stop codon (italicized). After PCR, the DNA fragments containing the integrase gene were cut with *Nde*I and *Kpn*I, and the DNA fragments containing the *lexA* gene were cut with *Kpn*I and *Bam*HI. The cleaved DNA fragments were purified with a Qiaex gel extraction kit (Qiagen) and ligated to pT7-7(His) plasmid DNA, previously cut with *Nde*I and *Bam*HI. Plasmid pT7-7(His) is derived from pT7-7, a T7 RNA polymerase-promoter system (46), and was prepared by inserting a double-stranded oligonucleotide (5′-TA*ATG*CATCACCATCACCATCACCA-3′ and 5′-TATGGTGATGGTGATGGTGATGCAT-3′) that contains an ATG initiation codon (italicized) and seven histidine codons (underlined) into the unique *Nde*I site of pT7-7.

To prepare a plasmid that contains a single specific binding site for LexA protein, a double-stranded oligonucleotide (5′-CAGGCCTGTATGAGCATACAGGTAC-3′ and 5′-CTGTATGCTCATACAGGCCTGGTAC-3′) containing the *recA* operator sequence (underlined) was inserted into the *Kpn*I site of a plasmid derived from pBluescript KSII+ (Stratagene), resulting in pBS-LA.

The sequences of all the PCR-amplified DNA fragments were verified by restriction analysis and the dideoxynucleotide chain termination method. Sequencing reactions were carried out with a modified T7 polymerase (Sequenase version 2.0; U.S. Biochemical) according to the manufacturer's specification.

Expression and purification of the fusion proteins. The DNA constructs were transformed into *E. coli* BL21(DE3). The cells were grown at 30°C. When the optical density at 600 nm was 0.8 to 1, 0.4 mM isopropyl-1-thio-β-D-galactopyranoside was added for expression induction, and the culture was grown for an additional 3 h.

(i) Purification in denaturing conditions. The cell pellet was resuspended in a buffer (5 ml of buffer per g of cells) containing 20 mM Tris-HCl (pH 8), 0.5 M NaCl, and 6 M guanidine-HCl (buffer A). The suspension was frozen and thawed, homogenized by stirring for 1 h at room temperature, and spun at 27,000 × *g* for 30 min at 4°C. The supernatant was passed twice over a Ni^{2+}-charged metal-chelating column (Qiagen) in the presence of 6 M guanidine-HCl at room temperature. Each column passage included a wash with buffer A, a second wash with buffer A plus 20 mM imidazole, and elution with a linear gradient from buffer A plus 20 mM imidazole to buffer A plus 500 mM imidazole. The fractions containing the protein were pooled and dialyzed in a stepwise manner against buffer B {25 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid [HEPES; pH 7.5], 1 mM EDTA, 10 mM dithiothreitol [DTT], 300 mM NaCl, 10% glycerol, 10 mM 3-[(3-cholamidopropyl)-dimethyl-ammonio]-1-propanesulfonate [CHAPS]} plus 1 M guanidine-HCl at 4°C. A 1.5-ml protein sample was then applied at 0.5 ml/min to a Superdex 75 (Pharmacia Biotech) column (about 100-ml resin bed volume) at 4°C. The fractions containing the protein were pooled and dialyzed against buffer B.

(ii) Purification in native conditions. The cell pellet was resuspended in a buffer containing (final concentrations) 20 mM HEPES (pH 7.5), 1 M NaCl, 10% glycerol, 5 mM 2-mercaptoethanol, 0.2 mM EDTA, 1 mM phenylmethylsulfonyl fluoride, 0.2 mg of lysozyme per ml, and 0.1% Nonidet P-40. The cell suspension was sonicated and centrifuged at 100,000 × *g* for 1 h at 4°C. The supernatant, after dialysis against buffer C (20 mM HEPES [pH 7.5], 1 M NaCl, 10% glycerol, 5 mM 2-mercaptoethanol, 0.1% Nonidet P-40), was incubated on ice for 2 h with the Ni-nitrilotriacetate resin. The resin was sequentially washed with buffer C, buffer C plus 10 mM imidazole, buffer C plus 50 mM imidazole, and buffer C plus 70 mM imidazole. The resin was then packed in a column, and the protein was eluted with a linear gradient from buffer C plus 70 mM imidazole to buffer C plus 500 mM imidazole. The fractions containing the protein were pooled, concentrated on a Centricon-10 column (Amicon), and dialyzed against the final buffer (20 mM HEPES [pH 7.5], 0.5 M NaCl, 20% glycerol, 0.1 mM EDTA, 1 mM DTT, 10 mM CHAPS). Protein concentrations were determined by the Bradford method (Bio-Rad), using bovine serum albumin (BSA) as a standard.

The wild-type integrase (WT IN) and fusion proteins IN1-288/LA, IN1-234/LA, and IN50-234/LA were purified in both native and denaturing conditions. For each protein, no difference in activity was observed when the protein was purified in either condition. Proteins IN50-234 and IN50-288/LA were purified under the native condition only, whereas proteins IN1-234, IN1-288/LABD, and IN1-234/LABD were purified under the denaturing condition only.

Footprinting analysis of DNA binding. pBS-LA plasmid DNA, which contains the LexA-binding sequence, was digested with *Bam*HI. The linearized DNA was labeled at the 5′ end with [γ-^{32}P]ATP by using T4 polynucleotide kinase and digested with *Pvu*II. The 311-bp single-end-labeled fragment containing the LexA-binding sequence was isolated from a 1.2% agarose gel with a Qiaex gel extraction kit (Qiagen). About 6 fmol (30,000 cpm) of the fragment was incubated with the protein at room temperature for 30 min in a buffer containing (final concentrations) 20 mM HEPES (pH 7.5), 10 mM DTT, 0.05% Nonidet P-40, 1.5 mM $CaCl_2$, 2.5 mM $MgCl_2$, 100 μg of BSA per ml, 2 μg of poly(dI-dC) per ml, and 50 mM NaCl. The samples were digested with 2 ng of DNase I per ml for 3 min at room temperature. The digestion was stopped by the addition of 18 mM EDTA, and the samples were deproteinized by phenol-chloroform extraction, ethanol precipitated in the presence of 10 μg of tRNA as a carrier, and resuspended in 5 μl of formamide–10 mM EDTA. After denaturation at 90°C for 3 min, the samples were analyzed by electrophoresis through a 5% denaturing polyacrylamide gel.

Integration assays. The 3′-end processing, 3′-end joining, and disintegration activities of the fusion proteins were assayed as previously described (9, 52). The following oligonucleotides (Operon Technologies, Inc.) were used as DNA substrates: T1 (16-mer), 5′-CAGCAACGCAAGCTTG-3′; T3 (30-mer), 5′-GTCGACCTGCAGCCCAAGCTTGCGTTGCTG-3′; V2 (21-mer), 5′-ACTGCTAGAGATTTTCCACAT-3′; V1/T2 (33-mer), 5′-ATGTGGAAAATCTCTAGCAGGCTGCAGGTCGAC-3′; C220 (21-mer), 5′-ATGTGGAAAATCTCTAGCAGT-3′; and B2-1 (19-mer), 5′-ATGTGGAAAATCTCTAGCA-3′. The oligonucleotides were purified by electrophoresis through a 15% denaturing polyacrylamide gel. Oligonucleotides T1, C220, and B2-1 were labeled at the 5′ end with [γ-^{32}P]ATP (6,000 Ci/mmol; Amersham) by using T4 polynucleotide kinase. The 3′-end processing and 3′-end joining substrate, which corresponds to the terminal 21 nucleotides of the U5 end of viral DNA, was prepared by annealing the labeled C220 strand with its complementary oligonucleotide V2. The preprocessed substrate, which resembles the viral U5 end after 3′-end processing, was prepared by annealing the labeled B2-1 strand with the V2 strand and was used to assay only the 3′-end joining activity. The substrate for assaying disintegration activity, the Y oligomer, was prepared by annealing the labeled T1 strand with oligonucleotides T3, V2, and V1/T2 (9). In a 20-μl volume, the DNA substrate (0.1 pmol) was incubated with the protein for 1 h at 37°C in the standard reaction buffer containing (final concentrations) 20 mM HEPES (pH 7.5), 10 mM DTT, 0.05% Nonidet P-40, and 10 mM $MnCl_2$. The reaction was stopped by the addition of 18 mM EDTA. The reaction products were heated at 90°C for 3 min before analysis by electrophoresis on 15% polyacrylamide gels with 7 M urea in Tris-borate-EDTA buffer.

Assays for distribution of integration sites. The donor DNA substrate used to assay the distribution of integration sites of the HIV integrase-LexA fusion proteins was the preprocessed U5 DNA substrate (B2-1/V2). The target DNA was plasmid pBS-LA (see above). The distribution of the integration sites was analyzed by the following assays.

(i) Agarose gel assay. pBS-LA was cleaved with *Mbo*II to generate multiple fragments ranging in size from 0.1 to 1 kbp (see Fig. 4). The DNA fragments (1 μg) were incubated with WT IN or with the fusion protein for 5 min at room temperature in the standard reaction buffer. The integration reaction was started by adding 15 nM preprocessed U5 substrate (B2-1/V2), labeled at the 5′ end of B2-1, and transferring the reaction to 37°C. After a 30-min incubation, the

reaction was stopped by adding 2 μl of 0.2 M EDTA (pH 8.0). The total reaction volume was 20 μl. The reaction product was mixed with a 1/6 volume of loading buffer (30% glycerol, 0.25% bromophenol blue, 0.25% xylene cyanol) and separated by electrophoresis on a 1.5% agarose gel in Tris-borate-EDTA buffer. After electrophoresis, the DNA fragments were visualized by ethidium bromide staining (0.5 μg/ml) and autoradiography.

(ii) PCR assay. The distribution and frequency of integration at individual DNA sites were also measured by a PCR-based assay performed as described previously (36). One microgram of plasmid pBS-LA was incubated with the protein at room temperature for 5 min in the standard reaction buffer. The integration reaction was started by adding 15 nM preprocessed U5 DNA (B2-1/V2) and incubating the sample at 37°C. After 30 or 60 min, the reaction was stopped by the addition of 15 mM EDTA (final concentration). The sample was extracted with phenol-chloroform, ethanol precipitated in the presence of 10 μg of tRNA, and washed with 70% ethanol. The pellet was resuspended in 50 μl of 10 mM Tris-HCl–1 mM EDTA (pH 7.5). A 5-μl aliquot of the reaction mixture was amplified for 25, 27, or 30 cycles of PCR (1 min at 94°C, 1 min at 55°C, and 2 min at 72°C). For analysis of the integration events occurring in the plus strand of the plasmid DNA, the PCR primers used were 0.2 μM unlabeled B2-1, 0.05 μM 5′-end-labeled B2-1, and 0.25 μM BS+ (5′-CATTAATGCAGCTGGCACGA-3′), which is complementary to the plus strand of the plasmid DNA and is located at 232 bp from the 3′ end of the LexA-binding sequence. For analysis of the integration events occurring in the minus strand, primer BS+ was replaced by primer BS− (5′-TAATACGACTCACTATAGGG-3′), which is complementary to the minus strand of the plasmid DNA and is located at 140 bp from the 3′ end of the LexA-binding sequence. The PCR was performed in a buffer containing (final concentrations) 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 0.001% (wt/vol) gelatin, 1.5 mM $MgCl_2$, 200 μM deoxynucleoside triphosphates, and 1 U of *Taq* polymerase (Perkin-Elmer Corp.) in a final volume of 20 μl. The labeled PCR products were analyzed on a denaturing 5% polyacrylamide gel and visualized by autoradiography.

RESULTS

Experimental design. The selection of integration sites was studied by fusing integrase to the *E. coli* LexA repressor, a sequence-specific DNA-binding protein. The LexA repressor of *E. coli* negatively regulates the transcription of about 20 SOS genes that are mostly involved in DNA repair, mutagenesis, DNA replication, and cell division (for reviews, see references 29 and 43). LexA protein was chosen for the fusion experiment because the structure and biochemical properties of the protein are well characterized. LexA protein contains two domains; the first 87 amino acids at the N terminus constitute the DNA-binding domain, and amino acid residues 88 to 202 constitute the dimerization domain (17, 42, 48). LexA protein binds specifically to a 16-bp DNA sequence that consists of two dyad symmetric half-sites of 8 bp each, starting with a highly conserved CTG trinucleotide that is followed by a less conserved but AT-rich 5-bp sequence (56). The sequence used in this study corresponds to the *recA* operator, a site that LexA binds with high affinity (28). Another reason for choosing LexA protein is that several studies have shown that the ability of LexA to bind to specific DNA sequences is retained after LexA is fused to various other proteins (5, 19, 41, 55).

The various fusion proteins constructed and analyzed in this study are shown in Fig. 1. The fusion protein consisting of full-length HIV-1 integrase fused to LexA (IN1-288/LA) serves as the prototype. Two fusion constructs, IN1-288/LABD and IN1-234/LABD, were prepared for determining whether fusion proteins containing only the DNA-binding domain of LexA were sufficient for altering target site selection. Since the central core of integrase contains the catalytic site (12) and the C terminus of integrase shows nonspecific DNA binding (15, 40, 53, 59), we prepared several fusion constructs that included various truncated forms of integrase, such as IN1-234/LA, IN50-288/LA, and IN50-234/LA. We were interested in whether the fusion proteins containing truncated integrase have an increased specificity toward the LexA-binding sequence in target site usage compared with those containing full-length integrase.

In vitro activities of the purified fusion proteins. All fusion proteins were first tested in the oligonucleotide-based assays for their abilities to mediate 3′-end processing, 3′-end joining, and disintegration (see Materials and Methods). Representative autoradiographs are shown in Fig. 2, and the results are summarized in Table 1. Fusing integrase with either full-length LexA or only the DNA-binding domain of LexA did not change appreciably the catalytic activities of integrase, and the two fusion proteins, IN1-288/LA and IN1-288/LABD, showed 3′-end processing and 3′-end joining activities similar to those of WT IN. For the 3′-end joining reaction, the patterns and the intensities of the recombinant products were similar among WT IN, IN1-288/LA, and IN1-288/LABD (Fig. 2A), indicating that fusion with LexA also did not alter the recognition by integrase of target DNA containing nonspecific sequences. Integrases containing various truncations and fusion proteins containing truncated integrase were inactive in 3′-end joining and 3′-end processing but retained disintegration activity (Fig. 2 and Table 1), similar to previous results obtained from mutational analysis of HIV-1 integrase (54). Although the truncated variants of integrase, either by themselves or fused with LexA, did not exhibit 3′-end joining activity in the oligonucleotide-based assays, the ability of each of these proteins to mediate 3′-end joining was demonstrated by a more sensitive PCR-based assay

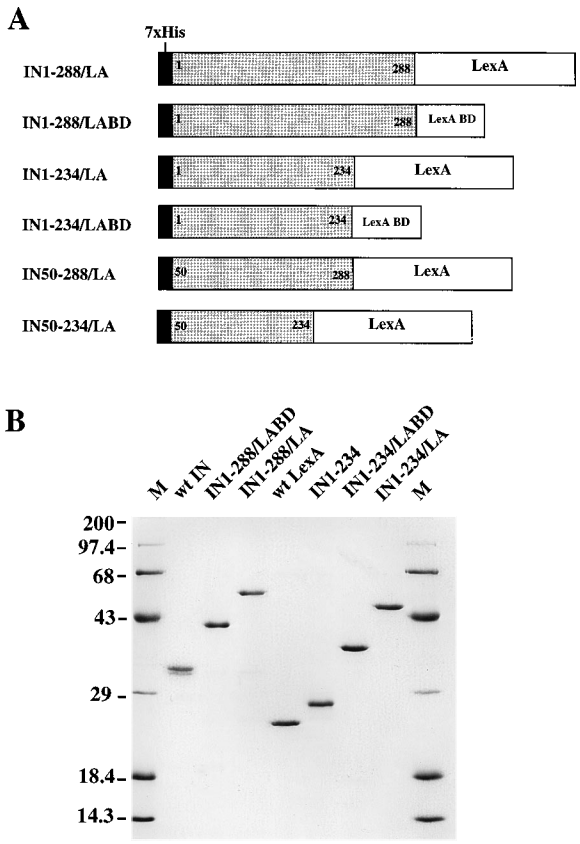

FIG. 1. (A) Primary structures of HIV-1 integrase-*E. coli* LexA fusion proteins. Open and stippled boxes represent peptides derived from HIV-1 integrase and LexA proteins, respectively. Filled boxes represent the seven consecutive histidine residues (7xHis) used for protein purification. The left and right ends of the boxes denote the amino and carboxy termini, respectively, of the fusion proteins. The numbers in the boxes correspond to the amino acid residues from the native protein included in each fusion protein. Full-length HIV-1 integrase and LexA have 288 and 202 amino acids, respectively. LexA, full-length LexA protein; LexA BD, DNA-binding domain (amino acid residues 1 to 87) of LexA. (B) Coomassie blue-stained sodium dodecyl sulfate (SDS)-polyacrylamide gel of various purified proteins. One microgram of each purified protein was run on an SDS–12% polyacrylamide gel. Lanes M contain the molecular weight standards (Gibco BRL); masses (in kilodaltons) are indicated on the left.

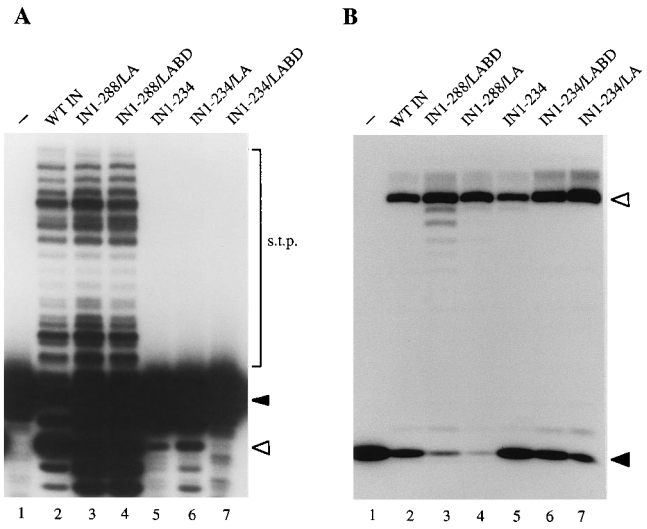

FIG. 2. Catalytic activities of HIV-1 integrase-LexA fusion proteins. (A) 3′-end processing and 3′-end joining activities. The reaction was carried out with 5 nM the U5 end oligonucleotide (C220/V2) and 100 nM the indicated proteins (lanes 2 to 7). The filled arrowhead indicates the position of the substrate (21-mer), and the open arrowhead designates the position of the 3′-end processing product (19-mer). s.t.p., strand transfer products. (B) Disintegration activity. The reaction was carried out with 5 nM the Y-oligomer substrate and 250 nM the indicated proteins (lanes 2 to 7). The filled arrowhead indicates the position of the 5′-end-labeled T1 strand of the Y-oligomer substrate, which migrated as a 16-nucleotide on the denaturing gel. The open arrowhead corresponds to the disintegration product (30-mer). In each panel, lane 1 represents a reaction done in the absence of protein.

(see below). IN1-186/LA did not display any catalytic activities (data not shown), and the protein was not studied further.

Fusing WT IN or truncated integrase to full-length LexA or only the DNA-binding domain of LexA increased the disintegration activity of the cognate protein (Fig. 2B). The mechanism for the stimulation is not clear. The stimulation may be related to an earlier observation that fusion of integrase of Rous sarcoma virus with various short peptides increased its activities (8).

The abilities of fusion proteins to recognize and bind specifically to a LexA-binding sequence were examined by DNase I footprinting analysis (Fig. 3). As expected from previous findings (15, 53, 59), WT IN and IN1-234 did not display any specific DNA binding on this DNA fragment, and the patterns were identical to that obtained in the absence of any protein (Fig. 3; compare lanes 3 and 7 with lanes 2 and 10). With the

TABLE 1. Summary of in vitro activities of HIV-1 integrase mutants and fusion proteins

Integrase derivative	Relative activity[a]		
	3′-end processing	3′-end joining	Disintegration
IN1-288/LA	++	++	+++
IN1-288/LABD	++	++	+++
IN1-234/LA	–	–[b]	+++
IN1-234/LABD	–	–[b]	+++
IN1-234	–	–[b]	+
IN50-288/LA	–	–[b]	++
IN50-234/LA	–	–[b]	++
IN50-234	–	–[b]	+

[a] Expressed by symbols representing the percentage of the activity of WT IN. +, 50% or less; ++, wild-type level of activity; +++, 150% or more; –, no activity.

[b] Although little or no 3′-end joining activity was observed in the oligonucleotide-based assay, strand transfer products were detected in the PCR-based assay (see Fig. 10).

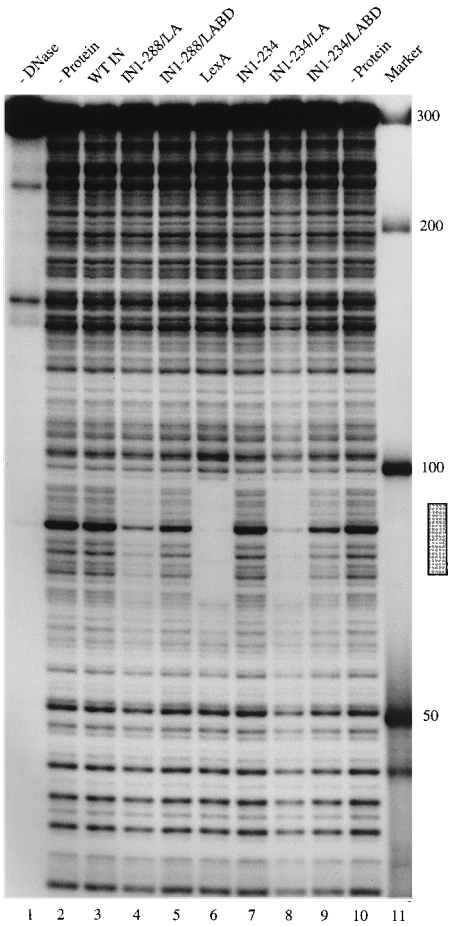

FIG. 3. Footprinting analysis of protein binding to a LexA recognition sequence. The DNA fragment (0.3 nM) was isolated after *Bam*HI and *Pvu*II digestion of pBS-LA, labeled at one 5′ end, and incubated with 100 nM the indicated proteins (lanes 3 to 9) for 30 min at room temperature. The samples were digested with DNase I (2 ng/ml) for 3 min at room temperature, and the digested products were separated on a denaturing polyacrylamide gel. Lane 1 represents the undigested, 5′-end-labeled DNA fragment. Lanes 2 and 10 represent reactions carried out in the absence of protein and subjected to DNase I treatment. Lane 11 contains size markers; lengths of DNA (in nucleotides) are indicated on the right. The location of the LexA-binding site is indicated by the stippled box.

wild-type LexA protein (Fig. 3, lane 6), a protected region of about 25 bp in size was observed, which is consistent with previous results (4, 23, 30). Protection of the LexA-binding sequence was also observed with the various fusion proteins (Fig. 3, lanes 4, 5, 8, and 9), providing direct evidence for sequence-specific DNA binding of these proteins. By calculating the amount of protein necessary to protect 50% of the sequence (2, 3), the dissociation constants of LexA, IN1-288/LA, IN1-288/LABD, IN1-234/LA, and IN1-234/LABD were estimated to be 2, 10, 250, 5, and 150 nM, respectively. The stronger protection displayed by fusion proteins containing full-length LexA (Fig. 3, lanes 4 and 8) than by fusion proteins containing only the DNA-binding domain of LexA (Fig. 3, lanes 5 and 9) suggests that the full-length LexA protein fused to the HIV-1 integrase is still able to dimerize, which provides a cooperative mode of binding to the operator as previously described (23). For IN1-288/LA and IN1-234/LA (Fig. 3, lanes 4 and 8), the size of protection was identical to that of wild-type LexA protein, suggesting that a LexA dimer component of the fusion protein is primarily responsible for DNA binding.

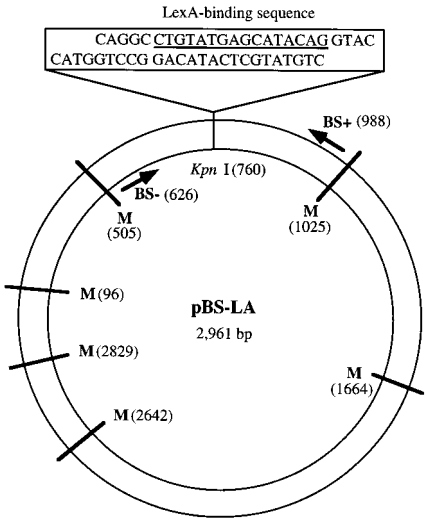

FIG. 4. DNA substrate for assaying distribution of integration sites. The LexA-binding sequence (underlined) was cloned into the *Kpn*I site of a plasmid derived from pBluescript KSII+. The resulting plasmid, pBS-LA, was digested with *Mbo*II to produce six fragments of different sizes (978, 639, 543, 409, 228, and 187 bp). The LexA-binding site is present in the 543-bp fragment. The arrows represent the primers used in PCR amplification of the integration products occurring in the plus or minus strand of the plasmid DNA. Primer BS+ is complementary to the plus strand of pBS-LA, whereas primer BS− is complementary to the minus strand. The numbers in parentheses denote the map positions of the sites for primer annealing and restriction enzyme cleavage. M, *Mbo*II.

Directed integration mediated by integrase-LexA fusion protein. An agarose gel assay was used to examine the distribution of integration sites of the integrase-LexA fusion proteins. The donor DNA for the agarose gel assay was labeled, preprocessed U5 DNA (B2-1/V2), and the target DNA was *Mbo*II-cleaved pBS-LA (Fig. 4), a plasmid that contains a LexA-binding site. *Mbo*II digestion of pBS-LA generates multiple DNA fragments with sizes ranging from 0.1 to 1 kbp. The fragment that contains the LexA-binding sequence is 543 bp in length (Fig. 4). Formation of recombinant products by integration of the labeled U5 DNA into target DNA was assayed by the appearance of labeled, high-molecular-weight DNA fragments. In the presence of WT IN, integration appeared to be random and occurred in all of the DNA fragments with similar frequencies (Fig. 5B, lanes 2 to 5). The integration frequency with WT IN increased at higher protein concentrations, but the relative intensities among the various DNA fragments remained the same. In contrast, integration of the U5 DNA by the fusion protein IN1-288/LA was unevenly distributed and showed a bias toward the DNA fragment containing the LexA-binding sequence (Fig. 5B, lane 7). In the presence of 2 pmol of fusion protein, the molar ratio between the DNA fragment containing the LexA-binding sequence and the IN1-288/LA dimer was about 1:1. The 543-bp LexA-containing DNA fragment was preferred approximately 14-fold over the other fragments (Fig. 5B, lanes 3 and 7). At higher concentrations of IN1-288/LA, the integration frequency increased but the bias became less apparent (Fig. 5B, lanes 8 and 9). In the reaction performed with 10 pmol of IN1-288/LA, the preference for the 543-bp fragment was approximately 4-fold (Fig. 5B, lanes 5 and 9). The result suggests that integration mediated by the integrase-LexA fusion protein was directed through specific DNA binding toward the fragment containing the LexA-binding sequence. The decrease in the selectivity at higher protein concentrations may be due to a saturation of binding of the LexA-binding site, which then caused the excess fusion protein to mediate integration randomly into other DNA fragments.

An identical experiment was carried out with IN1-288/LABD as the integration protein. The result obtained with IN1-288/LABD was similar to that obtained with IN1-288/LA. The distribution of integration sites of the fusion protein containing only the LexA-binding domain also exhibited a preference for the LexA-binding sequence, but the bias was approximately twofold less than that of IN1-288/LA (data not shown).

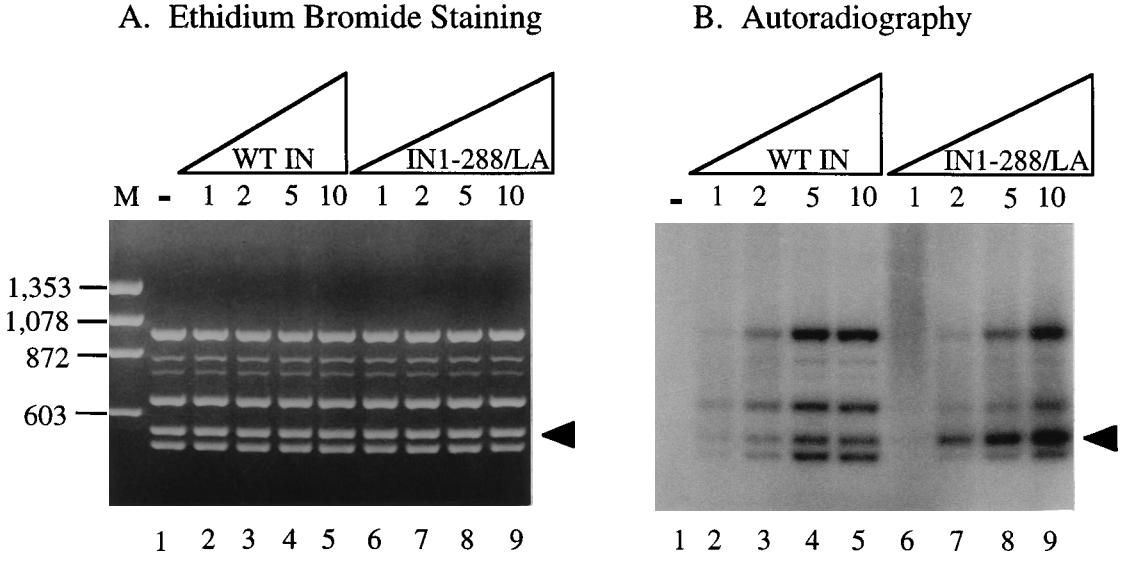

FIG. 5. Selective integration mediated by HIV integrase-*E. coli* LexA fusion protein. WT IN (lanes 2 to 5) or IN1-288/LA (lanes 6 to 9) at indicated amounts (in picomoles) was preincubated with 1 μg of *Mbo*II-cleaved pBS-LA for 5 min at room temperature before the reaction was started by adding 0.3 pmol of labeled preprocessed U5 DNA (B2-1/V2). The reaction products were separated by electrophoresis on a 1.5% agarose gel. Lane M corresponds to the molecular weight markers; sizes (in base pairs) are indicated on the left. Filled arrowheads indicate the positions of the DNA fragment containing the LexA-binding sequence. −, no protein. In panel A, the two faint bands that migrated slightly faster than the 872-bp marker are due to incomplete *Mbo*II digestion of the plasmid DNA. The frequency of integration mediated by WT IN or IN1-288/LA into the two smallest *Mbo*II-cleaved products, of 187 and 228 bp, was approximately threefold less than that into the 409-bp fragment (data not shown).

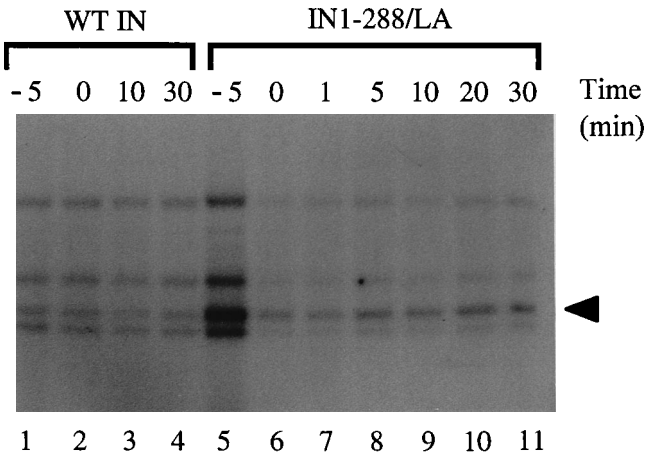

FIG. 6. Effect of preincubation of IN1-288/LA with target DNA. Two picomoles of WT IN (lanes 1 to 4) or IN1-288/LA (lanes 5 to 11) was preincubated with 1 μg of *Mbo*II-cleaved pBS-LA at room temperature for 0 (lanes 2 and 6), 1 (lane 7), 5 (lane 8), 10 (lanes 3 and 9), 20 (lane 10), or 30 (lanes 4 and 11) min before the addition of preprocessed U5 DNA. In lanes 1 and 5, the protein was preincubated at room temperature for 5 min with preprocessed U5 DNA before the reaction was started by adding target DNA. The arrowhead indicates the position of the fragment containing the LexA-binding sequence.

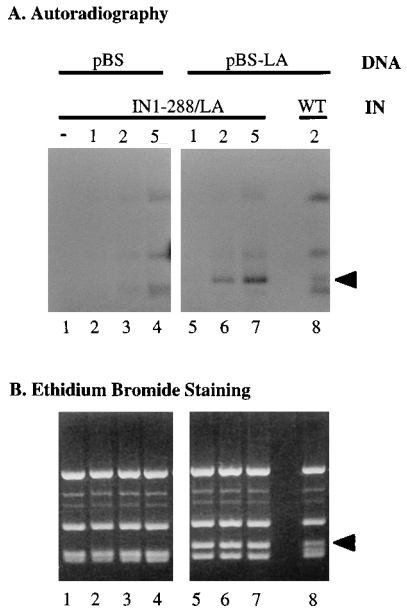

FIG. 7. Dependence on the LexA-binding sequence for selective integration. Integration of preprocessed U5 DNA was carried out by using WT IN (lane 8) or IN1-288/LA (lanes 1 to 7), with 1 μg of *Mbo*II-cleaved pBS (lanes 1 to 4) or *Mbo*II-cleaved pBS-LA (lanes 5 to 8) as the target DNA. The numbers above the lanes indicate the amounts of protein in picomoles. Arrowheads indicate positions of the fragment containing the LexA-binding sequence. In pBS, which has no LexA-binding sequence, the fragment corresponding to the 543-bp fragment of pBS-LA is 521 bp in length.

This result could be due to the lower binding affinity of IN1-288/LABD than of IN1-288/LA (Fig. 3). Our result is consistent with an earlier study showing that DNA binding by many LexA derivatives that contain the C-terminal dimerization domain is considerably higher than binding by fusions that lack it (19).

Because of the poor 3′-end joining activity of the truncated integrase-LexA fusion proteins (Fig. 2 and Table 1), the distribution of their integration sites was not determined in the agarose gel assay. Instead, the target site usage of these fusion proteins was examined by using a more sensitive PCR-based assay (see below).

We observed that the target site selection depended largely on whether the fusion protein was preincubated with the target DNA or the donor DNA (Fig. 6). The DNA fragment containing the LexA-binding sequence was preferred when the fusion protein was preincubated with the target DNA, although the time of preincubation was not critical (Fig. 6, lanes 6 to 11). In contrast, when the fusion protein was preincubated with the donor DNA, the integration events became more evenly distributed (Fig. 6, lane 5). In the case of the WT IN, no difference was observed regardless of whether the protein was preincubated with the target or donor DNA (Fig. 6, lanes 1 to 4). The result is consistent with the hypothesis that the preferred integration is mediated by the specific interaction between the fusion protein and the LexA-binding sequence and that such an interaction is promoted when the fusion protein is preincubated with the target DNA.

Directed integration by the fusion protein depends on the LexA-binding site and can be competed for by LexA protein. To confirm that the directed integration by the fusion protein relied on the presence of the LexA-binding sequence, we examined the distribution of integration sites into DNA fragments generated from *Mbo*II cleavage of the parental plasmid pBS, which contains no LexA-binding sequence. Under the identical reaction conditions and in the absence of the LexA-binding sequence in the target DNA, the IN1-288/LA fusion protein showed no bias in the frequency of integration (Fig. 7). The result indicates that the 543-bp fragment, except in the presence of the LexA-binding sequence, possessed no preferred sequence or DNA features that could have caused the directed integration.

A competition experiment was carried out to test the hypothesis that the directed integration observed with the fusion protein was mediated by its specific binding to the LexA-binding sequence (Fig. 8). Integration reactions mediated by a fixed concentration of WT IN or IN1-288/LA were done in the presence of different concentrations of LexA protein, and *Mbo*II-cleaved pBS-LA was used as the target DNA. In the presence of an increasing amount of LexA protein, the preferred integration mediated by IN1-288/LA into the DNA fragment containing the LexA-binding sequence correspondingly diminished, and the integration became more evenly distributed among all DNA fragments (Fig. 8B). The result is consistent with the model that LexA protein competes with the fusion protein for the LexA-binding site, resulting in free fusion protein that mediates random integration. Moreover, the LexA-bound DNA fragment, with the LexA-binding site being occupied, can no longer be specifically targeted. As a negative control, addition of LexA protein to the reaction containing WT IN had no effect on the distribution of integration sites (Fig. 8A). The unaltered usage of integration sites by WT IN and LexA protein also ruled out the possibility that the directed integration by the fusion protein could be an artifact resulting from DNA distortion induced by LexA protein binding.

Detailed analysis of integration sites by using the PCR-based assay. The usage of target sites can be assessed by using a PCR-based assay that has a much higher sensitivity and resolution than the agarose gel assay (36). The recombinant molecules were amplified by PCR, and the amplified products were resolved on a sequencing gel. Each band on the gel corresponds to an integration event at a given phosphodiester bond. The frequency of integration at a particular site and its exact position can be determined by the intensity of the band and by use of a sequencing ladder, respectively. Using the PCR assay, we compared WT IN and IN1-288/LA for the distribution and frequency of integration events around the LexA

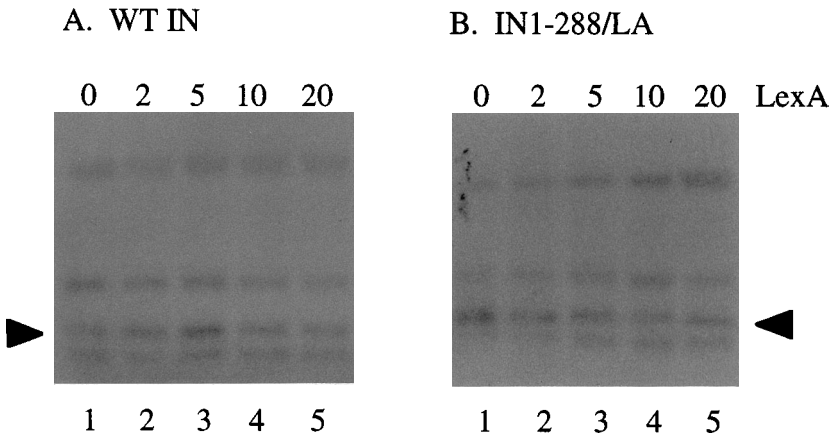

FIG. 8. Loss of selective integration by addition of *E. coli* LexA protein. Integration reactions were performed with 2 pmol of WT IN or IN1-288/LA in the presence of 0 to 20 pmol of LexA repressor. The LexA protein was preincubated with the target DNA for 5 min at room temperature before the reaction was started by adding WT IN or IN1-288/LA and 0.3 pmol of 5′-end-labeled U5 DNA. Arrowheads mark positions of the DNA fragment containing the LexA-binding site.

recognition sequence (Fig. 9). In the case of WT IN, with the LexA-binding site absent (pBS; Fig. 9A, lanes 1 to 5) or present (pBS-LA; Fig. 9B, lanes 1 to 5) in the target DNA, the distribution and intensity of the PCR-amplified products showed that most positions on the DNA could be used as target sites for integration, and there was a wide variation in integration frequency among the target sites. With the fusion protein IN1-288/LA, when the LexA-binding sequence was absent in the target DNA, the integration pattern was similar to that of WT IN (Fig. 9A, lanes 6 to 9). When the LexA-binding sequence was present in the target DNA, in contrast to WT IN, the LexA-binding region was not used as a target by the fusion protein, and a majority of the integration events instead occurred near the regions flanking the LexA-binding sequence (Fig. 9B, lanes 6 to 9). Concurrently, there was a notable decrease in the frequency of integration in the outlying region (30 bp or more) of the LexA-binding sequence. Several integration hot spots (arrowheads in Fig. 9B), located within 30 bp from the LexA-binding site, were found on the plus (Fig. 9) and minus (data not shown) strands of the target DNA. These hot spots were specific for the fusion protein and were not used as active target sites by the WT IN.

As a negative control, the integration reaction was carried out in the presence of a fixed amount of WT IN and various amounts of LexA protein (Fig. 9C). As the concentration of LexA protein increased in the reaction, there was a proportional decrease in the integration events occurring in the LexA-binding sequence. However, in contrast to the integration pattern observed with IN1-288/LA, there was no increase in integration in the regions flanking the LexA-binding sequence or decrease in integration in the outlying regions. The data provide support that the integration pattern of IN1-288/LA results from two components working in *cis* and not from a combined effect of two separate functions provided in *trans* by individual components.

An integration reaction using the PCR assay was also performed with the fusion protein IN1-288/LABD in order to examine possible differences in integration patterns between fusion proteins containing full-length LexA or only the DNA-binding domain of LexA. The integration pattern of IN1-288/LABD was similar to that of IN1-288/LA except that the pattern of IN1-288/LABD was less specific since there was more integration within the LexA-binding sequence as well as the outlying regions (data not shown). The result is consistent with the earlier findings from the agarose gel assay and the footprinting analysis (Fig. 3).

Target site usage of truncated integrase-LexA fusion proteins. The central core region of integrase contains the catalytic domain, and the C terminus of the protein is reported to bind nonspecific DNA. To determine the minimal domain required for the preferred integration and to test whether higher specificity could be achieved by using an integrase without the nonspecific DNA-binding domain, we examined the integration patterns of fusion proteins containing various truncations of integrase by the PCR assay (Fig. 10). The integration efficiencies of the truncated integrases, either by themselves or as fusion proteins, were approximately 100-fold lower than those of their full-length counterparts. Other than the poor efficiency, the integration patterns of the truncated integrases IN50-234 (Fig. 10, lane 2) and IN1-234 (data not shown) were surprisingly similar to that of WT IN (Fig. 9B). Likewise, the integration patterns of fusion proteins containing a truncated integrase, such as IN50-234/LA (Fig. 10, lane 3), IN50-288/LA (Fig. 10, lane 4), and IN1-234/LA (Fig. 10, lane 5), were similar to that of IN1-288/LA (Fig. 9B). The results showed that fusion proteins containing truncated integrase do not have any greater specificity than the fusion protein containing the full-length integrase. Furthermore, the similarity in integration patterns among the full-length protein and the various truncated integrases suggests that the protein domain responsible for target site recognition may reside in the central core region of integrase.

DISCUSSION

One salient feature of retroviral integration is that many sites on the host DNA can be used as targets (10, 58). We have demonstrated in vitro that one effective means of altering target site usage of integrase is by fusing integrase with a sequence-specific DNA-binding protein. In this study, we fused full-length or truncated HIV-1 integrase with full-length *E. coli* LexA or only the DNA-binding domain of the protein. The fusion proteins preferentially integrated viral DNA into target DNA containing the LexA-binding sequence. Analysis of the distribution and frequency of integration sites indicates that the fusion proteins first bind specifically to the LexA-binding sequence and then mediate integration in the nearby regions flanking the binding site. Our conclusion is supported by the following observations. (i) The preferred integration of the fusion proteins depended on the presence of the LexA protein component and was proportional to the binding affinities of the fusion proteins to the LexA-binding sequence. No preferred

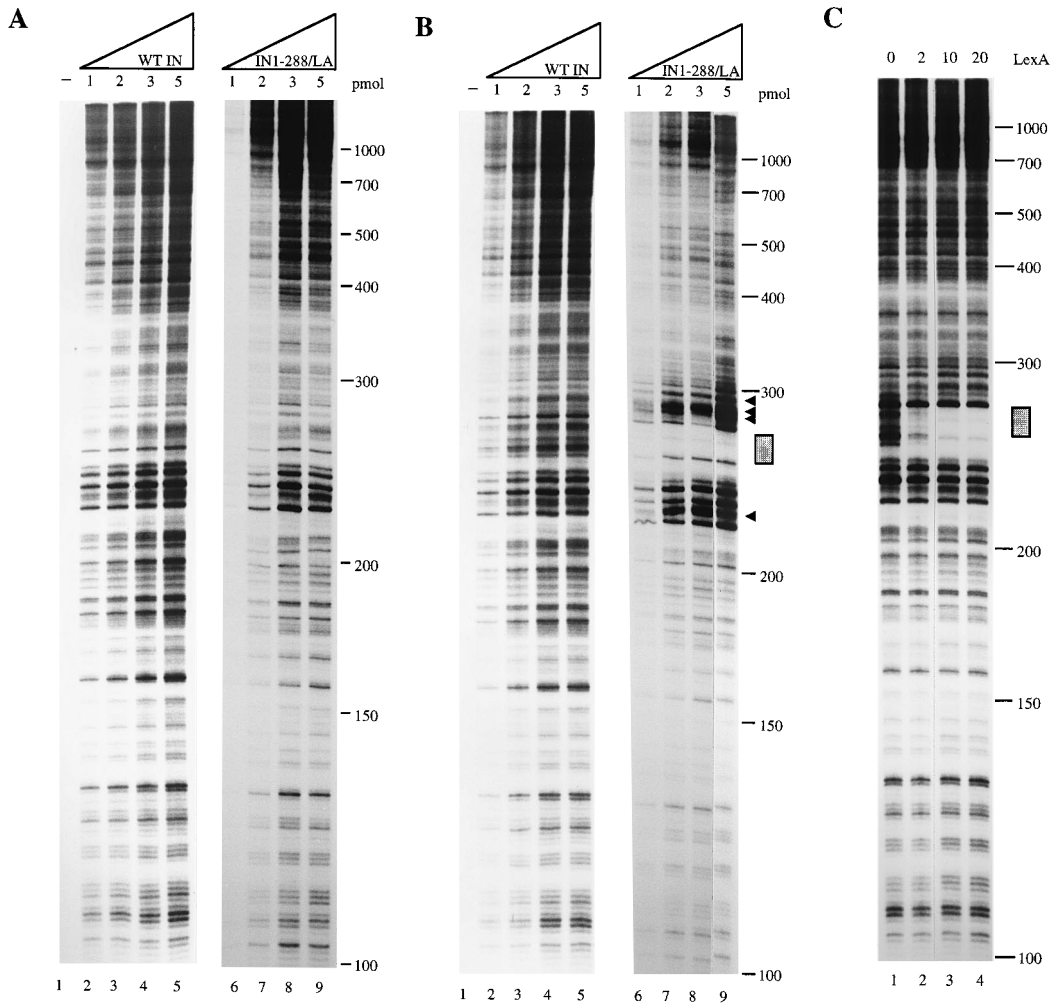

FIG. 9. Selection of target sites by WT IN and the fusion protein IN1-288/LA. (A) Target DNA without a LexA-binding sequence. One microgram of pBS DNA, which does not contain a LexA-binding sequence, was used as the target DNA and preincubated with the indicated concentrations of WT IN (lanes 2 to 5) or IN1-288/LA (lanes 6 to 9) for 5 min on ice. The integration reaction was started by adding 0.3 pmol of preprocessed U5 DNA and incubating the mixture at 37°C for 30 min. The reaction products were amplified by PCR using oligonucleotides B2-1 and BS+ as primers. Lane 1 is a negative control and represents the PCR products of an integration reaction carried out in the absence of WT IN or IN1-288/LA. Numbers on the right indicate lengths (in nucleotides) of size markers. (B) Target DNA containing a LexA-binding sequence. Experiments were performed identically to those described for panel A except that the target DNA used was from pBS-LA, which contains a LexA-binding sequence inserted at the *Kpn*I site (see Fig. 4). In comparison with lanes 6 to 8, lane 9 is of lighter exposure to show the banding pattern flanking the LexA-binding site. Arrowheads mark the integration hot spots specific for IN1-288/LA. The stippled box indicates the position of the LexA-binding site. (C) Selection of target sites by WT IN in the presence of LexA protein. One microgram of target DNA pBS-LA was preincubated with the indicated concentrations of LexA protein for 10 min at room temperature. WT IN (2 pmol) was then added, and the sample was preincubated for another 5 min at 4°C before the start of the integration reaction. Symbols have the same significance as in panel B.

integration was observed with WT IN or truncated HIV-1 integrases. (ii) The preferred integration depended on the presence of the LexA-binding sequence. In the absence of the LexA-binding sequence in the target DNA, the usage of target sites of fusion proteins was random and was identical to that of WT IN. In addition, preincubation of the target DNA with the fusion protein increased the integration specificity. (iii) The preferred integration was unique to the fusion proteins, and no preferred integration was observed when the reaction was performed with a mixture of WT IN and LexA protein.

Using the full-length integrase-LexA fusion protein, IN1-288/LA, we found that even though a majority of the integration was directed toward the DNA fragment containing the LexA-binding sequence, a considerable amount of integration occurred in fragments that contain no LexA-binding sequences. The nonspecific integration mediated by IN1-288/LA became more prominent at higher protein concentrations or when the fusion protein was not preincubated with the target DNA before the start of the reaction. We reasoned that the integrase component of IN1-288/LA presumably retains its functional target DNA-binding domain; the nonspecific integration, therefore, could be due to nonspecific DNA binding by the integrase component. This model could also explain the observations that the nonspecific integration was lowered by using subsaturating concentrations of IN1-288/LA or by prebinding the fusion protein with the LexA-binding sequence.

In an attempt to increase the integration specificity of the fusion protein, we constructed several fusion proteins that contain a truncated integrase. Special attention was paid to the C terminus of integrase, since this domain has been shown to bind DNA nonspecifically (15, 40, 53, 59). The close similarity of the integration patterns, determined by the PCR-based assay, of IN1-288/LA and the various truncated integrase-LexA fusion proteins indicates that no added specificity was achieved

by removing the N or C terminus of integrase. The result indicates that though the C terminus contributes to nonspecific DNA binding, it is unlikely to be involved in target site selection. Our result with respect to the integration pattern of the truncated integrases suggests that the integrase domain responsible for target site selection may reside in the central core (amino acid residues 50 to 234) of the protein. The result is consistent with a previous report showing that target site preference is altered by single amino acid substitutions of asparagine at position 120 within the central core domain of HIV-2 integrase (49).

One interesting finding is that the truncated integrases IN1-234 and IN50-234 showed a weak 3′-end joining activity when assayed by the sensitive PCR-based method; no 3′-end joining activity was detectable in the conventional in vitro assays. We also observed a weak 3′-end joining activity when we performed the same PCR assay with a D116N mutant, which contains an asparagine instead of the highly conserved aspartic acid at position 116 (data not shown). The weak 3′-end joining activity observed with the truncated integrases and the D116N mutant was not changed in the presence or absence of the N-terminal His tag (data not shown). The D116N mutant has been shown previously to be inactive in all known catalytic activities of integrase in the conventional assays (13, 26, 27, 49). Control experiments were carried out to confirm that the observed 3′-end joining activities of the truncated integrases and D116N mutant were not due to a contamination of the PCR (data not shown). The similarity among the mutant and wild-type integrases in the banding pattern on a sequencing gel further supports the view that the PCR-amplified products were not experimental artifacts and that the truncated integrases and D116N mutant indeed possess 3′-end joining activity. We think that this finding has important significance for in vivo experiments in which putatively integration-defective viruses are studied. Several studies have shown that viruses containing a mutation at D-116 of integrase, although not able to replicate, are positive for an indicator cell assay that requires the expression of Tat protein (1, 14, 57). The results are generally interpreted as evidence that unintegrated viral DNA may serve as a template for Tat expression. In light of the weak 3′-end joining activity of the D116N mutant, it is possible that viruses containing a D-116 mutation of integrase may be capable of forming a low level of proviruses, which may in turn produce sufficient Tat protein for the indicator cell assay.

Directed integration of viral DNA by a fusion protein consisting of the DNA-binding domain of λ repressor at the N terminus and HIV-1 integrase at the C terminus has been described previously (7). The λ repressor-integrase hybrid directs integration selectively into target DNA containing λ operators; many of the integration sites are near the operator on the same face of the B-DNA helix. For the integrase-LexA fusion proteins, most of the integration hot spots were also active integration sites for WT IN. The few new hot spots unique to the integrase-LexA fusion protein did not show any periodicity. The basis for the difference between the integration patterns of the two fusion proteins is not known. The difference may reflect the different modes of binding between λ repressor and LexA protein (11, 17, 37) or the fact that the DNA-binding domains are fused in different ways to integrase. Despite the difference in the distribution of integration sites, the two fusion proteins, made up of two different DNA-binding proteins placed at different termini of HIV-1 integrase, share the ability to target integration specifically to the cognate DNA sequences. Therefore, we believe that fusion proteins consisting of integrase and a sequence-specific DNA-binding protein can be of general use to achieve site-directed integration.

Retroviruses are now widely used as vectors for genetic engineering in higher eukaryotes and are considered to be promising vectors for gene therapy because of their natural aptitude for introducing foreign genes into cellular chromosomes (32). However, several features of current retroviral vectors, including the limited sizes of their genomes, their inability to infect nondividing cells, and their inability to target integration to a specific site (32, 45, 47), limit their usefulness in gene therapy. Studies to identify the determinants of target site selection may overcome this last problem.

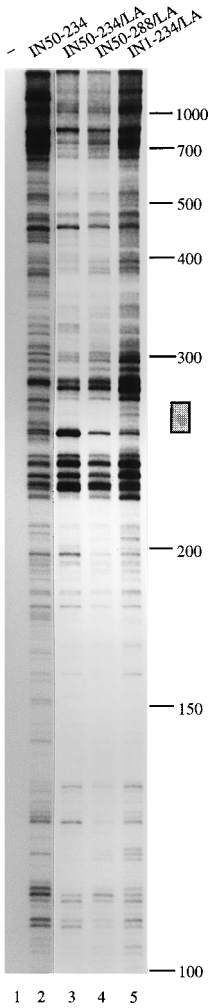

FIG. 10. Integration site usage of fusion proteins containing truncated integrase. The integration reaction was carried out for 1 h at 37°C in the presence of 250 nM the indicated protein. The recombinant products were amplified by PCR using oligonucleotides B2-1 and BS+ as primers. Twenty-seven cycles of PCR were performed for IN50-288/LA (lane 4) and IN1-234/LA (lane 5), and 30 cycles were performed for IN50-234 (lane 2) and IN50-234/LA (lane 3). Lane 1 represents an integration reaction performed in the absence of protein and subsequently amplified by 30 cycles of PCR as described earlier. Numbers and symbols on the right have the same significance as in Fig. 9.

ACKNOWLEDGMENTS

We thank Pat Brown, Yoshio Shibagaki, Jennifer Gerton, and Jean-François Mouscadet for helpful discussions, and we thank Janice Chow, TaiYun Roe, Pat Brown, and Fedrick Bushman for critical reading of the manuscript.

H.G. was supported by a postdoctoral fellowship from the Association de la Recherche contre le Cancer, France. This work was sup-

ported by grants from the Department of Energy (DE-FC03-87-ER60615) and National Institutes of Health (R01 CA68859).

REFERENCES

1. **Ansari-Lari, M. A., L. A. Donehower, and R. A. Gibbs.** 1995. Analysis of human immunodeficiency virus type 1 integrase mutants. Virology **211:**332–335.
2. **Brenowitz, M., D. Senear, E. Jamison, and D. Dalma-Weishausz.** 1993. Footprinting of nucleic acid-protein complexes, p. 1–43. *In* A. Revzin (ed.), Quantitative DNase I footprinting. Academic Press, Inc., New York.
3. **Brenowitz, M., D. F. Senear, M. A. Shea, and G. K. Ackers.** 1986. Quantitative DNase footprint titration: a method for studying protein-DNA interactions. Methods Enzymol. **130:**132–181.
4. **Brent, R., and M. Ptashne.** 1981. Mechanism of action of the lexA gene product. Proc. Natl. Acad. Sci. USA **78:**4204–4208.
5. **Brent, R., and M. Ptashne.** 1985. A eucaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. Cell **43:**729–736.
6. **Brown, P. O.** 1990. Integration of retroviral DNA. Curr. Top. Microbiol. Immunol. **157:**19–48.
7. **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.
8. **Bushman, F. D., and B. Wang.** 1994. Rous sarcoma virus integrase protein: mapping functions for catalysis and substrate binding. J. Virol. **68:**2215–2223.
9. **Chow, S. A., K. A. Vincent, V. Ellison, and P. O. Brown.** 1992. Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus. Science **255:**723–726.
10. **Craigie, R.** 1992. Hotspots and warm spots: integration specificity of retroelements. Trends Genet. **8:**187–190.
11. **Dumoulin, P., P. Oertel-Buchheit, M. Granger-Schnarr, and M. Schnarr.** 1993. Orientation of the LexA DNA-binding motif on operator DNA as inferred from cysteine-mediated phenyl azide crosslinking. Proc. Natl. Acad. Sci. USA **90:**2030–2034.
12. **Dyda, F., A. B. Hickman, T. M. Jenkins, A. Engelman, R. Craigie, and D. R. Davies.** 1995. Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. Science **266:**1981–1986.
13. **Engelman, A., and R. Craigie.** 1992. Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro. J. Virol. **66:**6363–6369.
14. **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.
15. **Engelman, A., A. B. Hickman, and R. Craigie.** 1994. The core and carboxy-terminal domains of the integrase protein of human immunodeficiency virus type 1 each contribute to nonspecific DNA binding. J. Virol. **68:**5911–5917.
16. **Fitzgerald, M. L., and D. P. Grandgenett.** 1994. Retroviral integration: in vitro host site selection by avian integrase. J. Virol. **68:**4314–4321.
17. **Fogh, R. H., G. Ottleben, H. Ruterjans, M. Schnarr, R. Boelens, and R. Kaptein.** 1994. Solution structure of the LexA repressor DNA binding domain determined by 1H NMR spectroscopy. EMBO J. **13:**3936–3944.
18. **Goff, S. P.** 1992. Genetics of retroviral integration. Annu. Rev. Genet. **26:**527–544.
19. **Golemis, E. A., and R. Brent.** 1992. Fused protein domains inhibit DNA binding by LexA. Mol. Cell. Biol. **12:**3006–3014.
20. **Grandgenett, D. P., R. B. Inman, A. C. Vora, and M. L. Fitzgerald.** 1993. Comparison of DNA binding and integration half-site selection by avian myeloblastosis virus integrase. J. Virol. **67:**2628–2636.
21. **Johnson, M. S., M. A. McClure, D. Feng, J. Gray, and R. F. Doolittle.** 1986. Computer analysis of retroviral pol genes: assignment of enzymatic functions to specific sequences and homologies with nonviral enzymes. Proc. Natl. Acad. Sci. USA **83:**7648–7652.
22. **Kalpana, G. V., S. Marmon, W. Wang, G. R. Crabtree, and S. P. Goff.** 1994. Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. Science **266:**2002–2006.
23. **Kim, B., and J. W. Little.** 1992. Dimerization of a specific DNA-binding protein on the DNA. Science **255:**203–206.
24. **Kirchner, J., C. M. Connolly, and S. B. Sandmeyer.** 1995. Requirement of RNA polymerase III transcription factors for in vitro position-specific integration of a retroviruslike element. Science **267:**1488–1491.
25. **Kitamura, Y., Y. M. Ha Lee, and J. M. Coffin.** 1992. Nonrandom integration of retroviral DNA *in vitro*: effect of CpG methylation. Proc. Natl. Acad. Sci. USA **89:**5532–5536.
26. **Kulkosky, J., K. S. Jones, R. A. Katz, J. P. G. Mack, and A. M. Skalka.** 1992. Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. Mol. Cell. Biol. **12:**2331–2338.
27. **Leavitt, A. D., L. Shiue, and H. E. Varmus.** 1993. Site-directed mutagenesis of HIV-1 integrase demonstrates differential effects on integrase functions in vitro. J. Biol. Chem. **268:**2113–2119.
28. **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.
29. **Little, J. W., and D. W. Mount.** 1982. The SOS regulatory system of *Escherichia coli*. Cell **29:**11–22.
30. **Little, J. W., D. W. Mount, and C. R. Yanisch-Perron.** 1981. Purified LexA protein is a repressor of the RecA and LexA genes. Proc. Natl. Acad. Sci. USA **78:**4199–4203.
31. **Muller, H. P., and H. E. Varmus.** 1994. DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes. EMBO J. **13:**4704–4714.
32. **Mulligan, R. C.** 1993. The basic science of gene therapy. Science **260:**926–932.
33. **Pruss, D., F. D. Bushman, and A. P. Wolffe.** 1994. Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core. Proc. Natl. Acad. Sci. USA **91:**5913–5917.
34. **Pruss, D., R. Reeves, F. D. Bushman, and A. P. Wolffe.** 1994. The influence of DNA and nucleosome structure on integration events directed by HIV integrase. J. Biol. Chem. **269:**25031–25041.
35. **Pryciak, P. M., A. Sil, and H. E. Varmus.** 1992. Retroviral integration into minichromosomes in vitro. EMBO J. **11:**291–303.
36. **Pryciak, P. M., and H. E. Varmus.** 1992. Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. Cell **69:**769–780.
37. **Ptashne, M.** 1992. A genetic switch. Cell Press and Blackwell, Cambridge, Mass.
38. **Rohdewohld, H., H. Weiher, W. Reik, R. Jaenisch, and M. Breindl.** 1987. Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. J. Virol. **61:**336–343.
39. **Sandmeyer, S. B., L. J. Hansen, and D. L. Chalker.** 1990. Integration specificity of retrotransposons and retroviruses. Annu. Rev. Genet. **24:**491–518.
40. **Schauer, M., and A. Billich.** 1992. The N-terminal region of HIV-1 integrase is required for integration activity, but not for DNA-binding. Biochem. Biophys. Res. Commun. **185:**874–880.
41. **Schmidt-Dorr, T., P. Oertel-Buchheit, C. Pernelle, L. Bracco, M. Schnarr, and M. Granger-Schnarr.** 1991. Construction, purification, and characterization of a hybrid protein comprising the DNA binding domain of the LexA repressor and the Jun leucine zipper: a circular dichroism and mutagenesis study. Biochemistry **30:**9657–9664.
42. **Schnarr, M., M. Granger-Schnarr, S. Hurstel, and J. Pouyet.** 1988. The carboxy-terminal domain of the LexA repressor oligomerises essentially as the entire protein. FEBS Lett. **234:**56–60.
43. **Schnarr, M., P. Oertel-Buchheit, M. Kazmaier, and M. Granger-Schnarr.** 1991. DNA binding properties of the LexA repressor. Biochimie **73:**423–431.
44. **Shibagaki, Y., and S. A. Chow.** Unpublished data.
45. **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.
46. **Tabor, S., and C. C. Richardson.** 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. Proc. Natl. Acad. Sci. USA **82:**1074–1078.
47. **Temin, H. M.** 1990. Safety considerations in somatic gene therapy of human disease with retrovirus vectors. Hum. Gene Ther. **1:**111–123.
48. **Thliveris, A. T., and D. W. Mount.** 1992. Genetic identification of the DNA binding domain of *Escherichia coli* LexA protein. Proc. Natl. Acad. Sci. USA **89:**4500–4504.
49. **van Gent, D. C., A. A. M. Oude Groeneger, and R. H. A. Plasterk.** 1992. Mutational analysis of the integrase protein of human immunodeficiency virus type 2. Proc. Natl. Acad. Sci. USA **89:**9598–9602.
50. **Varmus, H. E., and P. O. Brown.** 1989. Retroviruses, p. 53–108. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
51. **Vijaya, S., D. L. Steffen, and H. L. Robinson.** 1986. Acceptor sites for retroviral integrations map near DNase I-hypersensitive sites in chromatin. J. Virol. **60:**683–692.
52. **Vincent, K. A., V. Ellison, S. A. Chow, and P. O. Brown.** 1993. Characterization of human immunodeficiency virus type 1 integrase expressed in *Escherichia coli* and analysis of variants with amino-terminal mutations. J. Virol. **67:**425–437.
53. **Vink, C., A. A. M. Oude Groeneger, and R. H. A. Plasterk.** 1993. Identification of the catalytic and DNA-binding region of the human immunodeficiency virus type 1 integrase protein. Nucleic Acids Res. **21:**1419–1425.
54. **Vink, C., and R. H. A. Plasterk.** 1993. The human immunodeficiency virus integrase protein. Trends Genet. **9:**433–437.
55. **Wang, H., and D. J. Stillman.** 1993. Transcriptional repression in *Saccharomyces cerevisiae* by a SIN3-LexA fusion protein. Mol. Cell. Biol. **13:**1805–1814.
56. **Wertman, K. F., and D. W. Mount.** 1985. Nucleotide sequence binding specificity of the LexA repressor of *Escherichia coli* K-12. J. Bacteriol. **163:**376–384.
57. **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.
58. **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.
59. **Woerner, A. M., M. Klutch, J. G. Levin, and C. J. Markus-Sekura.** 1992. Localization of DNA binding activity of HIV-1 integrase to the C-terminal half of the protein. AIDS Res. Hum. Retroviruses **8:**2433–2437.