Rous Sarcoma Virus Integrase Protein: Mapping Functions for Catalysis and Substrate Binding

FREDERIC D. BUSHMAN* AND BEIBEI WANG

Infectious Disease Laboratory, Salk Institute, La Jolla, California 92037

Received 7 October 1993/Accepted 13 December 1993

Rous sarcoma virus (RSV), like all retroviruses, encodes an integrase protein that is responsible for covalently joining the reverse-transcribed viral DNA to host DNA. We have probed the organization of functions within RSV integrase by constructing mutant derivatives and assaying their activities *in vitro*. We find that deletion derivatives lacking the amino-terminal 53 amino acids, which contain the conserved H-X₍₃₋₇₎-H-X₍₂₃₋₃₂₎-C-X₍₂₎-C (HHCC) Zn²⁺-binding motif, are greatly impaired in their ability to carry out two reactions characteristic of integrase proteins: specific cleavage of the viral DNA termini and DNA strand transfer. Deletion mutants lacking the carboxyl-terminal 69 amino acids are also unable to carry out these reactions. However, all deletion mutants that retain the central domain are capable of carrying out disintegration, an *in vitro* reversal of the normal DNA strand transfer reaction, indicating that the catalytic center probably lies within this central region. Another conserved motif, D-X₍₃₉₋₅₈₎-D-X₍₃₅₎-E, is found in this central domain. These findings with RSV integrase closely parallel previous findings with human immunodeficiency virus integrase, indicating that a modular catalytic domain is a general feature of this family of proteins. Surprisingly, and unlike results obtained so far with human immunodeficiency virus integrase, efficient strand transfer activity can be restored to a mutant RSV integrase lacking the amino-terminal HHCC domain by fusion to various short peptides. Furthermore, these fusion proteins retain the substrate specificity of RSV integrase. These data support a model in which the integrase activities required for strand transfer *in vitro*, including substrate recognition, multimerization, and catalysis, all lie primarily outside the amino-terminal HHCC domain.

In order to complete the early stages of replication, a retrovirus must synthesize a DNA copy of the viral RNA genome and integrate that copy into a chromosome of the host. After the completion of reverse transcription and prior to integration, the blunt ends of the linear viral cDNA are cleaved so as to remove two nucleotides from each 3' end. The recessed 3' ends are then joined to protruding 5' ends of staggered breaks in the target DNA. The resulting integration intermediate, which contains a single-stranded gap at each end, is then repaired to yield an integrated provirus. For a recent review, see reference 14.

The DNA-cutting and -joining reactions involved in the synthesis of the integration intermediate are carried out by the virus-encoded integrase protein. Previous work has established that purified integrase is capable of cleaving model DNA substrates to remove two nucleotides and then joining the recessed 3' ends to target DNAs *in vitro* (3, 5, 7, 18, 20, 28, 34, 35). In addition, integrase can carry out an *in vitro* reversal of the integration reaction, named disintegration, in which a branched DNA structure resembling an integration product is converted into two molecules resembling the initial viral and target DNAs (6) (Fig. 1).

The human immunodeficiency virus (HIV) integrase can be subdivided into three regions. Studies of the function of mixed multimers composed of HIV integrase mutants revealed that each of these regions can act independently in certain settings (Fig. 2) (9, 31). The three regions are as follows.

(i) **Central catalytic domain.** The central region of HIV integrase, roughly amino acids 50 to 212, was found to be

relatively resistant to proteolytic digestion (10), as expected for a stably folded protein domain. In addition, refolding studies revealed that HIV IN⁵⁰⁻²¹² was capable of independent folding. A purified fragment containing this central region can carry out disintegration *in vitro* (4, 33). Since the disintegration reaction likely involves chemistry similar to that of the normal terminal cleavage and strand transfer reactions, the central region is inferred to contain the active site for all the activities of integrase. This part of the protein contains an amino acid sequence motif, D-X₍₃₉₋₅₈₎-D-X₍₃₅₎-E (D,D-35-E motif), that is found in the integrase proteins of retroviruses and retrotransposons and also in the transposase proteins of the IS3 family of bacterial transposons (12, 22, 26). Two further observations support the view that the central domain comprises the catalytic center. First, point mutants in the conserved residues of the central domain diminish all the enzymatic activities of integrase in parallel (8, 10, 22, 24, 30). Second, both the terminal cleavage and strand transfer reactions have been found to take place by mechanisms that probably involve a single chemical step (11), as expected if both are carried out by a single active site.

(ii) **Amino-terminal Zn²⁺-binding domain.** The amino-terminal 50 amino acids also contain a conserved sequence motif, H-X₍₃₋₇₎-H-X₍₂₃₋₃₂₎-C-X₍₂₎-C (HHCC motif), that is found in the integrase proteins of retroviruses and retrotransposons (15). The amino-terminal domain of HIV integrase binds Zn²⁺, both by itself and in the context of the complete protein, and the conserved His and Cys residues are required for efficient Zn²⁺ binding (2, 4). No independent activities of the HIV IN¹⁻⁵⁰ peptide bound to Zn²⁺ have yet been detected, although derivatives of HIV integrase, Rous sarcoma virus (RSV) integrase, and Moloney murine leukemia virus integrase containing alterations in these conserved sequences have

* Corresponding author. Mailing address: Infectious Disease Laboratory, The Salk Institute, 10010 N. Torrey Pines Rd., La Jolla, CA 92037. Phone: (619) 453-4100, ext. 630. Fax: (619) 554-0341.

FIG. 1. Reactions catalyzed by integrase proteins in vitro. DNAs are shown as thin lines, DNA 5' ends are drawn as small circles, and a ^{32}P -labeled 5' end is drawn as the filled circle. DNA substrates resembling one end of the unintegrated linear viral DNA can be cleaved to remove two nucleotides from one 3' end (terminal cleavage), and the recessed 3' end can then be joined to the 5' end of a break made in the target DNA. An apparent reversal of the strand transfer reaction (disintegration) can also be catalyzed by integrase. For the disintegration reactions presented in this paper, the short DNA strand at the lower right as drawn was labeled with ^{32}P .

been found to be impaired in terminal cleavage and strand transfer (2, 8, 10, 17, 21, 24, 30, 32).

(iii) **Carboxyl-terminal region.** The carboxyl-terminal region is least similar among different integrases and contains no obvious amino acid sequence motifs. Both the amino- and carboxyl-terminal regions are required for normal terminal cleavage and strand transfer, but their roles in these reactions are unclear.

In the course of other studies, we found to our surprise that substantial activity was present in a deletion derivative of RSV integrase lacking the amino-terminal HHCC motif. To explore this observation further, we constructed a set of deletions of RSV integrase and fusions of these derivatives to new sequences. Derivatives of RSV integrase containing residues 49 to 286 fused to various short peptides (initially added to aid in purification or detection) were capable of detectable terminal cleavage and near-wild-type levels of strand transfer in buffers containing Mn^{2+} . The short fusion peptides were required for high activity, but several different peptides provided the complementing function. From these observations we infer that the amino-terminal domain is not playing a central role in integration in vitro and that the remainder of the protein, residues 49 to 286, contains the determinants that direct multimerization, carry out catalysis, and recognize the viral and target DNAs. As expected from this view, the deletion-fusion derivatives retain

FIG. 2. Diagrams of RSV and HIV integrases (IN), illustrating conserved amino acid sequence motifs. The numbers below each rectangle correspond to the number in the amino acid sequence. The letters above indicate the locations of the conserved HHCC and D,D-35-E motifs. The edges of probable protein domains of HIV integrase are marked with the vertical lines.

the specificity of RSV integrase when compared with HIV integrase on RSV and HIV substrates. The central D,D-35-E domain of RSV integrase, like that of HIV, is capable of catalyzing the disintegration reaction in vitro, indicating that in RSV integrase also the catalytic domain is a separable functional unit.

MATERIALS AND METHODS

DNA manipulation. Derivatives of RSV integrase were prepared by constructing plasmids encoding the modified proteins. Procedures for manipulating DNA molecules were essentially as described previously (27). A plasmid encoding RSV integrase residues 1 to 286 (p32 form) fused to a His Tag was constructed as follows. DNA encoding residues 1 to 286 was amplified from pATV8 (19), a DNA plasmid containing an infectious RSV provirus, by PCR using primers that attached an *NdeI* restriction site to DNA encoding the amino terminus and a *BclI* site to DNA encoding the carboxyl terminus. Amplification products were purified, cleaved with *NdeI* and *BclI*, and ligated to pET15b (Novagen) cleaved with *NdeI* and *BamHI*. These manipulations resulted in attachment of a 20-amino-acid His Tag (HT in designations; sequence, MGSSHHHHHHSSGLVPRGSH) encoded by pET15b in frame at the amino terminus of the RSV integrase coding region, yielding pFB254. To make (HT)RSV IN⁴⁹⁻²⁸⁶, pFB254 was digested with *NdeI* and *AvrII*, and a double-stranded synthetic linker DNA was ligated to this backbone, yielding pBW5. This synthetic linker DNA restored sequences encoding RSV integrase residues 49 to 53 and contained ends complementary to the *NdeI* and *AvrII* ends of pFB254. Plasmids encoding (HT)RSV IN¹⁻²¹⁷ and (HT)RSV IN⁴⁹⁻²¹⁷ were created by amplifying the appropriate DNA sequences from pATV8 so as to attach *NdeI* and *BclI* sites at each end and then ligating these fragments to pET15b digested with *NdeI* and *BamHI*, yielding pFB267 and pFB266. A plasmid encoding (HT)RSV IN⁴⁹⁻¹⁷⁰ was created by digesting pBW5 with *BamHI* and *Bpu1102I* and then ligating this backbone DNA to a duplex oligonucleotide that (i) restored codons 169 and 170, (ii) supplied a stop codon, and (iii) supplied ends complementary to the *BamHI* and *Bpu1102I* sites (yielding plasmid pFB272). A plasmid encoding RSV IN⁴⁹⁻²⁸⁶(HT) (HT sequence at the 3' end) was created by amplifying the region encoding residues 49 to 286 in pATV8 with primers that supplied *NdeI* and *HindIII* sites at each end and then ligating this fragment to pET21b (Novagen) cleaved with *NdeI* and *HindIII*, yielding pFB271. The sequence of the carboxyl-terminal His Tag is KLAAALEHHHHHH. A plasmid encod-

ing (T7)RSV IN^{49–286}(HT) was constructed by amplifying the region encoding residues 49 to 286 in pATV8 with primers that linked an *Eco*RI site to the region encoding the amino terminus and a *Hind*III site to the region encoding the carboxyl terminus. This DNA fragment was ligated to pET23b (Novagen) cleaved with *Eco*RI and *Hind*III, yielding pFB270. The sequence of the T7 Tag is MASMTGGGQMGDPNS. The structures of all DNA constructions were confirmed by restriction mapping and sequencing across the junctions between vector and insert DNA.

Protein purification. Derivatives of RSV integrase bearing the His Tag were purified essentially as described previously (Novagen) (4). Plasmids encoding the deletions were introduced into *Escherichia coli* BL21(DE3), and protein synthesis was induced as described previously (Novagen) (4). Cells from induced cultures were collected by centrifugation and resuspended in 1 × binding buffer (0.5 M NaCl, 20 mM Tris [pH 7.9], 5 mM imidazole). Cells were lysed by (i) freezing in liquid N₂ and thawing at 37°C, (ii) treatment with lysozyme (0.2 mg/ml) for 30 min on ice, and (iii) sonication. The lysate was then centrifuged for 30 min at 44,000 × *g*. The soluble fraction was filtered through a 0.45-μm filter and applied to a chelating fast-flow Sepharose (Pharmacia) column charged with NiSO₄. Washing and elution were carried out as described previously (Novagen). The eluate was then dialyzed against 1 M NaCl–20 mM HEPES (*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid) (pH 7.5) to reduce the imidazole concentration. In the cases of (HT)RSV IN^{49–286}, (HT)RSV IN^{49–170}, and RSV^{49–286}(HT), the dialysate was repressed over an NiSO₄ column and then dialyzed against 0.5 M NaCl–20 mM HEPES (pH 7.5)–0.1 mM EDTA–1 mM dithiothreitol–20% glycerol. In the cases of (HT)RSV IN^{1–217}, (HT)RSV IN^{49–217}, and (T7)RSV IN^{49–286}(HT), the majority of the protein precipitated during dialysis against 1 M NaCl–20 mM HEPES (pH 7.5). The precipitate could be efficiently resolubilized by extraction with 100 mM NaCl–20 mM HEPES (pH 7.5)–1 mM dithiothreitol–0.1 mM EDTA–1% Nonidet P-40. These extracted preparations were then used directly in activity assays. Protein concentrations were determined from UV spectra or by electrophoresis and comparison with protein standards of known concentrations. Purified RSV integrase lacking the His Tag was the gift of Ira Palmer and Paul Wingfield, to whom we are grateful.

Cleavage of (HT)RSV IN^{49–286} with thrombin was carried out in 200 mM NaCl–20 mM Tris (pH 8.4)–2.5 mM CaCl₂–10% glycerol in a final volume of 100 μl. One unit of thrombin (Novagen) was used per 7.2 μg of protein. Cleavage was carried out for 4 h at room temperature. Amino acid sequencing revealed that (HT)RSV IN^{49–286} lacked the amino-terminal methionine. Amino acid sequencing of the thrombin cleavage product revealed that cleavage occurred not between the R and G residues in the Tag but between R-53 and G-54 of RSV integrase. Amino acid sequencing was kindly carried out by Wolfgang Fischer.

Integration assays. The terminal cleavage substrate consisted of two hybridized oligonucleotides, 5'CTACAAGAG TATTGCATAAGACTACATT3' (FB141) and its complement (FB142). This sequence matches that of the U3 terminus of the unintegrated RSV DNA. The precleaved substrate consisted of FB142 hybridized to an oligonucleotide identical to FB141 except that the two 3' residues were absent (FB163). The disintegration substrate consisted of four oligonucleotides, 5'GAGTATTGCATAAGACTACAGGGGCTATGG CGTCC3' (FB174), 5'AATGTAGTCTTATGCAATACTC3' (FB175), 5'GAAAGCGACCGCGCC3' (FB133), and 5'GGA CGCCATAGCCCCGCGCGGTCGCTTTC3' (FB134). Annealing of these four oligonucleotides produced the Y-shaped

disintegration substrate (Fig. 1, bottom). Substrates were labeled by treatment of FB133, FB141, or FB163 with [γ -³²P]ATP and T4 polynucleotide kinase prior to hybridization. The disintegration reaction converted FB133 (15 bases) to a 30-base product.

Assay mixtures contained 5 mM MnCl₂ or MgCl₂, 25 mM HEPES (pH 7.5), 20 mM β -mercaptoethanol, 100 μg of bovine serum albumin per ml, and 10% glycerol. The final NaCl concentration (indicated in the legend to each figure) was largely determined by the amount of NaCl brought in with the most dilute enzyme. The derivatives of RSV integrase, like RSV integrase itself (13), were not very sensitive to changes in salt concentration between 40 and 125 mM (data not shown). The concentrations of integrase derivatives and substrates are indicated in the figure legends. Enzyme was added last to each assay, and reaction mixtures were incubated for 1 h at 37°C. Reactions were stopped by addition of sequencing gel loading dye containing excess EDTA, heated at 95°C for 3 to 5 min, and loaded onto 15% polyacrylamide DNA sequencing-type gels. Reaction products were visualized by autoradiography, and representative reactions were quantitated by counting Cherenkov emissions after excision of radioactive bands from the gels.

RESULTS

Deletion mutants of RSV integrase. To probe the functions of the various parts of RSV integrase, a series of deletion derivatives was created and analyzed. In an effort to make deletions of RSV integrase containing intact domains, deletions were constructed with breakpoints bearing the same relationship to the conserved amino acid sequence motifs as in HIV integrase (Fig. 2) (4). The first 50 amino acids in both RSV and HIV contain the HHCC motif in a roughly similar position, probably within a folded protein domain. A deletion of RSV integrase was constructed in which the first 48 amino acids were removed, generating (HT)RSV IN^{49–286}. Mutants are denoted RSV IN with the amino acids preserved in superscript, peptides fused to RSV IN are in parentheses, and HT denotes the His Tag sequence added to facilitate purification. In an attempt to make a stable central domain fragment, a stop codon was placed 50 amino acids carboxyl terminal of the last clearly conserved residue of the D,D-35-E motif, isoleucine 167 of RSV integrase, as in our most active derivative of the HIV central domain (HIV IN^{50–212}). This deletion was made in the context of complete RSV integrase, yielding (HT)RSV IN^{1–217}, and in the context of the amino-terminal truncation, yielding (HT)RSV IN^{49–217}. A further carboxyl-terminal deletion, (HT)RSV IN^{49–170}, was also constructed.

Purified mutant proteins and wild-type RSV integrase were analyzed for their relative activities in vitro on oligonucleotide substrates designed to monitor terminal cleavage, strand transfer, and disintegration. In addition, since RSV integrase is active in reactions containing either Mn²⁺ or Mg²⁺ (20), separate assays were conducted in the presence of each metal. (HT)RSV IN^{49–286}, the deletion retaining the His Tag, displayed a surprising range of activities, which are described separately below. For the experiments presented in this section, the His Tag was removed by cleavage with thrombin, yielding RSV IN^{54–286} (Fig. 3).

In reactions in vitro containing Mg²⁺, purified RSV integrase removes two nucleotides from one 3' end of a substrate mimicking one end of the unintegrated linear viral DNA (Fig. 4a, lane 1). About 5% of the starting substrate was converted to the –2 product in 1 h at 37°C. None of the deletion mutants displayed detectable specific cleavage (Fig. 4a, lanes 2 to 4). In

FIG. 3. Thrombin cleavage of (HT)RSV IN⁴⁹⁻²⁸⁶. Lane 1, (HT)RSV IN⁴⁹⁻²⁸⁶ incubated in the absence of thrombin; lane 2, (HT)RSV IN⁴⁹⁻²⁸⁶ incubated in the presence of thrombin. Reactions were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis, and proteins were visualized by staining with Coomassie blue. The mobilities of molecular weight standards (thousands) are as indicated. The two bands of higher molecular weight in lane 2 have the expected mobilities of thrombin (molecular weight, 33,000) and human serum albumin (molecular weight, 66,000; present in the thrombin preparation as a stabilizing agent).

reactions containing substrates lacking the two terminal 3' nucleotides (precleaved substrates), integration products are seen with wild-type integrase (Fig. 4b, lane 1; 2% of the substrate converted to product) but not with the deletion mutants (Fig. 4b, lanes 2 to 4). The strand transfer product appears as a ladder of bands because many DNA sequences can serve as target sites for DNA strand transfer, and integra-

tion into different positions yields product strands of different lengths. Wild-type integrase was capable of efficient disintegration in Mg²⁺ (Fig. 4c, lane 1; 20% of substrate converted to product). In this reaction, the top right strand of the target part of the disintegration substrate as drawn at the bottom of Fig. 1 becomes joined to the remainder of the top strand, converting the labeled 15-base oligonucleotide to a 30-base product. Reactions containing RSV IN⁵⁴⁻²⁸⁶ yielded low and somewhat variable amounts of disintegration product (Fig. 4c, lane 2; 0.4% of starting material converted to product in this experiment). Reactions containing (HT)RSV IN¹⁻²¹⁷ and (HT)RSV IN⁴⁹⁻²¹⁷ yielded extremely low levels of disintegration product (Fig. 4c, lanes 3 and 4, respectively, <0.2% of starting material converted to product) detectable only after very long exposures of the autoradiograms.

In reactions containing Mn²⁺, the nuclease activity of RSV integrase is much more prominent and less specific (see Fig. 5, 6, and 7) (20). This activity is probably not due to contaminating nucleases, since RSV integrase and derivatives purified to near homogeneity (i) by conventional chromatography, (ii) from a soluble fraction by the His Tag method, or (iii) from an insoluble fraction by the His Tag method show the same nuclease pattern (see Fig. 5, 6, and 7; and data not shown).

Cleavage at the expected -2 position is abundant in reactions containing wild-type RSV integrase and 5 mM MnCl₂ (Fig. 5a, lane 1; 20% of substrate converted to the -2 product) (Table 1). Strand transfer products can also be seen as a ladder of high-molecular-weight bands. Cleavage at the -2 position is detectable in reactions containing the deletion mutants but is much less prominent than with the wild type (Fig. 5a, lanes 2 to 4; 0.8 to 1.5% of substrate converted to the -2 product).

Reactions containing wild-type RSV integrase and the pre-cleaved substrate yielded prominent strand transfer products in the presence of 5 mM MnCl₂ (Fig. 5b, lane 1). Strand transfer products were also detected in reactions containing

FIG. 4. Activities of RSV integrase and deletion derivatives in the presence of 5 mM MgCl₂. Proteins used in each test reaction are as indicated above the lanes. (A) Analysis of terminal cleavage products. The terminal cleavage product is marked by the arrow. (B) Analysis of strand transfer products. The strand transfer products are marked by the bracket. (C) Analysis of disintegration products. The disintegration product is marked by the arrow. Each reaction mixture contained 18 pmol of integrase or derivative, 0.4 pmol of substrate, and 40 mM NaCl.

FIG. 5. Activities of RSV integrase and deletion derivatives in the presence of 5 mM $MnCl_2$. (A) Analysis of terminal cleavage products. (B) Analysis of strand transfer products. (C) Analysis of disintegration products. Markings are as in Fig. 4. Reaction mixtures for panels A and B contained 18 pmol of integrase or derivative and 40 mM NaCl; those for panel C contained 125 mM NaCl and 3 pmol of integrase or derivative. Each reaction mixture contained 0.4 pmol of substrate.

RSV IN⁵⁴⁻²⁸⁶ and the precleaved substrate (Fig. 5b, lane 2) but were much less than with the wild type (0.7 versus 7% of substrate converted to product). Strand transfer products were undetectable in reactions containing the other deletion mutants.

Reactions containing wild-type RSV integrase produced abundant disintegration product in the presence of 5 mM $MnCl_2$ (Fig. 5c, lane 1; 50% of substrate converted to product). Reactions containing RSV IN⁵⁴⁻²⁸⁶ also yielded abundant disintegration product (Fig. 5c, lane 2; 30% of substrate converted to product). Reactions containing (HT)RSV IN¹⁻²¹⁷ and (HT)RSV IN⁴⁹⁻²¹⁷ yielded somewhat less disintegration product (Fig. 5c, lanes 3 and 4, respectively; 5 and 4% of substrate converted to product, respectively). (HT)RSV IN⁴⁹⁻¹⁷⁰ yielded a very small amount of disintegration product, which was detectable only after very long autoradiographic exposures (Fig. 5c, lane 5, and data not shown; less than 0.7% of substrate converted to product). These data indicate that deletion derivatives of RSV integrase containing residues 54 to

217 alone are capable of disintegration, although residues 218 to 286 are required for maximal activity, while the full protein is required for efficient terminal cleavage and strand transfer (Table 1).

Activities of RSV IN⁴⁹⁻²⁸⁶ fused to various peptides. In the course of analyzing the deletion mutants described above, we found that (HT)RSV IN⁴⁹⁻²⁸⁶ displayed a range of activities that was unexpected on the basis of our previous experience with HIV integrase. In reactions containing Mn^{2+} , RSV IN⁵⁴⁻²⁸⁶ displayed only weak cleavage at position -2, while the derivative containing the His Tag, (HT)RSV IN⁴⁹⁻²⁸⁶, surprisingly displayed much higher levels (Fig. 6a, lanes 1 and 2, respectively; about 3% versus about 16% of substrate converted to the -2 product) (Table 2). Both proteins also yielded prominent cleavage at position -3. The cutting activity of (HT)RSV IN⁴⁹⁻²⁸⁶ is not fully wild type, since the wild type yields more of the -2 product than the -3 product, while the deletion-fusion mutant yields about equal quantities of each.

Most strikingly, in reactions containing the precleaved substrate and (HT)RSV IN⁴⁹⁻²⁸⁶, wild-type levels of strand transfer products were produced (Fig. 6b, lane 2; 8% of substrate converted to product). RSV IN⁵⁴⁻²⁸⁶, in contrast, displayed much weaker activity (Fig. 6b, lane 1, and Fig. 5b, lane 2; 0.7 and 1.8% of substrate converted to product, respectively).

What function does the amino-terminal His Tag perform? To probe this issue, we constructed two further deletion-fusion mutants. To determine whether the location of the His Tag relative to positions 49 to 286 was important and to facilitate further studies, we constructed RSV IN⁴⁹⁻²⁸⁶(HT), in which the His Tag is transferred to the carboxyl terminus. Reactions containing RSV IN⁴⁹⁻²⁸⁶(HT) displayed -2 cleavage activity greater than that with RSV IN⁵⁴⁻²⁸⁶, but the ratio of cleavage at position -2 to that at -3 was lower than with (HT)RSV IN⁴⁹⁻²⁸⁶ (Fig. 6a, lane 3). The level of strand transfer in

TABLE 1. Activities of RSV integrase and deletion derivatives on RSV substrates

Integrase derivative	% of wild-type activity ^a		
	Terminal cleavage	Strand transfer	Disintegration
RSV integrase	+++	+++	+++
RSV IN ⁵⁴⁻²⁸⁶	+	++	+++
(HT)RSV IN ¹⁻²¹⁷	+	—	++
(HT)RSV IN ⁴⁹⁻²¹⁷	+	—	+
(HT)RSV IN ⁴⁹⁻¹⁷⁰	+	—	—

^a +++, 40 to 100%; ++, 10 to 40%; +, 1 to 10%; —, not measurable above background. Reactions were conducted in the presence of 5 mM $MnCl_2$.

FIG. 6. Activities of deletion-fusion derivatives of RSV integrase in the presence of 5 mM $MnCl_2$. Proteins used in each test reaction are as indicated above the lanes. (A) Analysis of terminal cleavage products. (B) Analysis of strand transfer products. (C) Analysis of disintegration products. Markings are as in Fig. 4. Each reaction mixture contained 125 mM NaCl, 0.1 pmol of substrate, and 6 pmol of integrase or mutant derivative.

reactions containing RSV IN⁴⁹⁻²⁸⁶(HT) was well below the wild-type level seen with (HT)RSV IN⁴⁹⁻²⁸⁶ (Fig. 6b, lane 3; 2% of substrate converted to product), indicating that the location of the His Tag does influence strand transfer. The slight differences in activity between RSV IN⁵⁴⁻²⁸⁶ and RSV IN⁴⁹⁻²⁸⁶(HT) could be due to either the presence of the carboxyl-terminal His Tag or the five-amino-acid difference at the amino terminus generated by the thrombin cleavage.

Is the specific amino acid sequence of the His Tag important for its ability to substitute for the HHCC domain, or will any peptide sequence provide this activity? To examine this issue, an epitope tag derived from a 12-amino-acid sequence in gene 10 of phage T7 was fused to the amino terminus of RSV IN⁴⁹⁻²⁸⁶(HT). This mutant, (T7)RSV IN⁴⁹⁻²⁸⁶(HT), displayed -2 cleavage activity comparable to that of RSV IN⁴⁹⁻²⁸⁶(HT) (Fig. 6a, lane 4), although for unclear reasons the ratio of -2 to -3 cleavage was somewhat lower. Importantly, the levels of strand transfer in reactions containing (T7)RSV IN⁴⁹⁻²⁸⁶(HT) were higher than with RSV IN⁴⁹⁻²⁸⁶(HT) or RSV IN⁵⁴⁻²⁸⁶

(Fig. 6b, lane 4; 3.5% of substrate converted to product). Evidently the sequence of the peptide fused to the amino terminus is not critical for stimulating strand transfer.

All of the deletion-fusion mutants displayed disintegration activity in the presence of Mn^{2+} (Fig. 6c; 23 to 30% of substrate converted to product) (Table 2) and Mg^{2+} (data not shown), indicating that the catalytic domain is not grossly misfolded in the less active mutants. None of the deletion-fusion mutants displayed detectable specific cleavage or strand transfer in reactions containing Mg^{2+} (data not shown).

Activities of deletion-fusion mutants on HIV substrates. To assess the specificity of (HT)RSV IN⁴⁹⁻²⁸⁶, the activities on HIV substrates were compared with the activities of HIV integrase and RSV integrase (Table 3). Mn^{2+} was chosen as the cofactor for these experiments because HIV integrase displays only extremely weak activity in the presence of Mg^{2+} . RSV integrase retained some terminal cleavage activity on an HIV substrate, as reported previously (20), but HIV integrase displayed greater specificity and efficiency (Fig. 7a, lanes 1 and 2; 8% of substrate converted to product for HIV integrase

TABLE 2. Activities of deletion-fusion mutants on RSV substrates

Integrase derivative	% of wild-type activity ^a		
	Terminal cleavage ^b	Strand transfer	Disintegration
RSV IN ⁵⁴⁻²⁸⁶	+	++	+++
(HT)RSV IN ⁴⁹⁻²⁸⁶	+++	+++	+++
RSV IN ⁴⁹⁻²⁸⁶ (HT)	+++	++	+++
(T7)RSV IN ⁴⁹⁻²⁸⁶ (HT)	+++	+++	+++

^a Results are expressed as the percentage of the activity of RSV integrase. Symbols are as in Table 1. Reactions were conducted in the presence of 5 mM $MnCl_2$.

^b Although the level of -2 cleavage was within 40 to 100% of the wild-type level, a lower ratio of cleavage at position -2 to that at -3 was observed with the deletion-fusion mutants.

TABLE 3. Activities of HIV integrase, RSV integrase, and deletion-fusion mutants on HIV substrates

Integrase derivative	% of wild-type activity ^a		
	Terminal cleavage	Strand transfer	Disintegration
HIV integrase	+++	+++	+++
RSV integrase	++	+++	+++
RSV IN ⁵⁴⁻²⁸⁶	+	++	++
(HT)RSV IN ⁴⁹⁻²⁸⁶	++	+++	+++

^a Results are expressed as the percentage of the activity of HIV integrase. +++, 40 to 180%; ++, 10 to 40%; +, 1 to 10%. Reactions were conducted in the presence of 5 mM $MnCl_2$.

FIG. 7. Activities of HIV integrase, RSV integrase, and derivatives of RSV integrase on HIV substrate DNAs in the presence of 5 mM $MnCl_2$. Proteins used in each test reaction are as indicated above the lanes. (A) Analysis of terminal cleavage products. (B) Analysis of strand transfer products. (C) Analysis of disintegration products. Labels are as in Fig. 4. Each reaction mixture contained 50 mM NaCl, 0.8 pmol of substrate, and 15 pmol of integrase or mutant derivative.

versus 3% for RSV integrase). (HT)RSV IN⁴⁹⁻²⁸⁶ also showed weak cleavage at the -2 position (Fig. 7a, lane 4; 1% converted to product) and abundant cleavage at other positions in the substrate. RSV IN⁵⁴⁻²⁸⁶ displayed still weaker -2 cleavage activity (Fig. 7a, lane 3; 0.4% of substrate converted to product).

In reactions containing precleaved HIV substrates, RSV integrase and (HT)RSV IN⁴⁹⁻²⁸⁶ produced high levels of strand transfer product (Fig. 7b, lanes 2 and 4; 5 and 3% of substrate converted to product), as did HIV integrase (Fig. 7b, lane 1; 3% converted to product). Reactions with RSV IN⁵⁴⁻²⁸⁶ yielded much less product (Fig. 7b, lane 3; 0.4%). The pattern of bands in the integration product generated by RSV integrase was very different from that with HIV integrase, indicating that the selection of the target site is determined by the integrase and not the target DNA. This conclusion was also reached in previous studies comparing Moloney murine leukemia virus integrase and HIV integrase (23). The patterns of bands in the integration products with RSV integrase and (HT)RSV IN⁴⁹⁻²⁸⁶ were indistinguishable, indicating that target site specificity is similar in each.

HIV integrase, RSV integrase, and (HT)RSV IN⁴⁹⁻²⁸⁶ displayed robust activity on an HIV disintegration substrate (Fig. 7c, lanes 1, 2, and 4; 10, 6, and 7% of substrate converted to product, respectively), indicating that the sequence requirements in the viral DNA part of the disintegration substrate may not be very strict. In support of this view, we also find that HIV integrase is as active as RSV integrase on an RSV disintegration substrate (data not shown). The sequence of the target-like part of the HIV disintegration substrate was identical to that of the RSV disintegration substrate. RSV IN⁵⁴⁻²⁸⁶ displayed slightly less activity (Table 3) (Fig. 7c, lane 3; 3% of substrate converted to product).

DISCUSSION

In this paper, we present studies on the activities in vitro of deletion derivatives of RSV integrase. We find that a surprising range of activities is preserved in mutants that lack the amino-terminal HHCC domain, as long as short peptides are fused to the remainder of integrase. Such deletion-fusion mutants display wild-type levels of strand transfer and modest levels of specific cleavage in the presence of 5 mM $MnCl_2$. This observation was unexpected, since similar deletion mutants of HIV integrase are much more impaired in the terminal cleavage and strand transfer activities. On the basis of these observations, we propose that the amino-terminal HHCC domain is not strictly required for substrate binding, forming the active oligomer (9, 16, 31), or catalysis. We also find that the central region, containing the D,D-35-E sequence motif, can carry out disintegration when explanted from the rest of the protein, indicating that this region probably contains the catalytic center. A similar central fragment of HIV integrase is also capable of disintegration, indicating that a modular catalytic domain may be a common feature of the integrase family (4, 33).

Although the results with deletion-fusion mutants of RSV integrase differed from results with HIV integrase, the functions of the HHCC and D,D-35-E domains need not be qualitatively different in the two cases. In the case of HIV, weak stimulation of strand transfer by the attachment of a His Tag at the amino terminus has also been observed. HIV IN⁵⁰⁻²⁸⁸ lacking the His Tag was incapable of cleavage and joining (unpublished data), whereas a mutant with a His Tag attached to the amino terminus retained about 2% of the wild-type activity (4, 9). In the case of RSV, RSV IN⁴⁹⁻²⁸⁶(HT) retains about 25% of the wild-type strand transfer activity with Mn^{2+} , (T7)RSV IN⁴⁹⁻²⁸⁶(HT) retains 45% of the activity, and

(HT)RSV IN⁴⁹⁻²⁸⁶ displays 90 to 120% of the activity. Thus, strand transfer activity is greater in both systems when a short peptide is attached to the amino terminus of an HHCC-deleted mutant.

What role does the HHCC domain play in integration *in vitro*, and what roles do the complementing peptides play in the deletion-fusion mutants? Three models seem plausible.

(i) The HHCC domain may contribute to binding either the viral DNA or target DNA, a proposal consistent with the observation that Zn²⁺-binding domains often bind DNA (1). It could be argued that the His Tag provides a substitute metal-binding domain, since it contains six His residues and is known to be able to bind metal. Furthermore, the weakly positively charged His residues might make electrostatic contacts with the DNA phosphate backbone. However, the T7 Tag, which also exhibits complementing activity, has no net charge, and there is no reason to think that the T7 Tag can bind metal. Furthermore, the DNA substrate specificities of wild-type RSV integrase and the deletion-fusion mutants are similar for both viral DNA and target DNAs, a result that might not be expected if the amino-terminal peptides were major contributors to DNA binding.

(ii) Perhaps the HHCC domain stabilizes the folding of the rest of the protein, and the two complementing peptides exert a similar stabilizing influence. This model has the advantage of explaining how such different amino-terminal peptides can exert their stimulatory effects.

(iii) Perhaps the HHCC domain is required for proper orientation of monomers within the active multimer. For example, the HHCC domain or amino-terminal peptides might fill a cavity in the complex that when present allows the monomers to slip into an inactive arrangement.

Models ii and iii need not be strictly distinct, since altered conformations of each monomer could influence the conformation of the multimer and vice versa. In a recent study of RNase H of HIV type 1, a His Tag was found to be required for activity (29); in this case, the added peptide could have promoted either substrate binding or proper folding. In the case of RSV integrase, some combination of models ii and iii seems most likely to explain the complementing activity of the two peptides and, by extension, the role of the HHCC domain in the wild-type protein.

Why do the deletion-fusion mutants carry out strand transfer with Mn²⁺ but not Mg²⁺? Although the reason for the difference is unclear, the wider chemical bonding potential of Mn²⁺ (see, e.g., reference 25) may partially compensate for functional deficits of the deletion-fusion proteins by promoting interactions that are weaker or impossible with Mg²⁺. Different conformation of the DNA substrates in the presence of the two metals may also influence activity. Mg²⁺ is the more likely metal cofactor *in vivo*, and so the complementation by short peptides in the deletion-fusion mutants should be viewed as only partial. However, there is no reason to think that the role of the HHCC domain in the presence of Mn²⁺, is qualitatively different from that in the presence of Mg²⁺, so the simplest model is that the HHCC domain is playing a relatively nonspecific role in both settings.

The view that the HHCC domain plays a secondary role in integration reactions *in vitro* does not imply that the HHCC domain is unimportant *in vivo*. The evolutionary conservation of this domain establishes its significance. In integration reactions *in vitro*, it seems most likely that the HHCC domain serves as an architectural element that helps organize the activities of the carboxyl-terminal functions. In the presence of Mg²⁺, the likely *in vivo* cofactor, this function is indispensable.

Whether the HHCC domain is involved in additional functions in the context of the complete virus remains to be seen.

ACKNOWLEDGMENTS

We thank Janet Leatherwood Bushman, Marc Sitbon, Didier Trono, and members of our laboratory for comments on the manuscript.

This work was supported by NIH grant 1R01AI34786-0104. F.D.B. is a Special Fellow of the Leukemia Society of America.

REFERENCES

1. Berg, J. M. 1990. Zinc finger domains: hypothesis and current knowledge. *Annu. Rev. Biophys. Biophys. Chem.* **19**:405-421.
2. Burke, C. J., G. Sanyal, M. W. Bruner, J. A. Ryan, R. L. LaFemina, H. L. Robbins, A. S. Zeff, C. R. Middaugh, and M. G. Cordingley. 1992. Structural implications of spectroscopic characterization of a putative zinc finger peptide from HIV-1 integrase. *J. Biol. Chem.* **267**:9639-9644.
3. Bushman, F. D., and R. Craigie. 1991. Activities of human immunodeficiency virus (HIV) *in vitro*: specific cleavage and integration of HIV DNA. *Proc. Natl. Acad. Sci. USA* **88**:1339-1343.
4. Bushman, F. D., A. Engelman, I. Palmer, P. Wingfield, and R. Craigie. 1993. Domains of the integrase protein of human immunodeficiency virus type 1 responsible for polynucleotidyl transfer and zinc binding. *Proc. Natl. Acad. Sci. USA* **90**:3428-3432.
5. Bushman, F. D., T. Fujiwara, and R. Craigie. 1990. Retroviral DNA integration directed by HIV integration protein *in vitro*. *Science* **249**:1555-1558.
6. 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.
7. Craigie, R., T. Fujiwara, and F. Bushman. 1990. The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration *in vitro*. *Cell* **62**:829-837.
8. Drellich, M. R., R. Wilhelm, and J. Mous. 1992. Identification of amino acid residues critical for endonuclease and integration activities of HIV-1 IN protein *in vitro*. *Virology* **188**:459-468.
9. Engelman, A., F. D. Bushman, and R. Craigie. 1993. Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex. *EMBO J.* **12**:3269-3275.
10. Engelman, A., and R. Craigie. 1992. Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function *in vitro*. *J. Virol.* **66**:6361-6369.
11. Engelman, A., K. Mizuuchi, and R. Craigie. 1991. HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. *Cell* **67**:1211-1221.
12. Fayet, O., P. Ramond, P. Polard, M. F. Prere, and M. Chandler. 1990. Functional similarities between the IS3 family of bacterial insertion elements? *Mol. Microbiol.* **4**:1771-1777.
13. Fitzgerald, M. L., A. C. Vora, W. G. Zeh, and D. P. Grandgenett. 1992. Concerted integration of viral DNA termini by purified avian myeloblastosis virus integrase. *J. Virol.* **66**:6257-6263.
14. Goff, S. P. 1992. Genetics of retroviral integration. *Annu. Rev. Genet.* **26**:527-544.
15. Johnson, M. S., M. A. McClure, D.-F. Feng, J. Gray, and R. F. Doolittle. 1986. Computer analysis of retroviral pol genes: assignment of enzymatic functions to specific sequences and homologies with nonretroviral enzymes. *Proc. Natl. Acad. Sci. USA* **83**:7648-7652.
16. Jones, K. S., J. Coleman, G. W. Merkel, T. M. Laue, and A. M. Skalka. 1992. Retroviral integrase functions as a multimer and can turn over catalytically. *J. Biol. Chem.* **267**:16037-16040.
17. Jonsson, C. B., and M. J. Roth. 1993. Role of the His-Cys finger of Moloney murine leukemia virus integrase protein in integration and disintegration. *J. Virol.* **67**:5562-5571.
18. Katz, R. A., G. Merkel, J. Kulkosky, J. Leis, and A. M. Skalka. 1990. The avian retroviral IN protein is both necessary and sufficient for integrative recombination *in vitro*. *Cell* **63**:87-95.
19. Katz, R. A., C. A. Omer, J. H. Weis, A. Mitsialis, A. J. Faras, and R. V. Guntaka. 1982. Restriction endonuclease and nucleotide sequence analysis of molecularly cloned unintegrated avian tumor

- virus DNA: structure of large terminal repeats in circle junctions. *J. Virol.* **42**:346–351.
20. **Katzman, M., R. A. Katz, A. M. Skalka, and J. Leis.** 1989. The avian retroviral integration protein cleaves the terminal sequences of linear viral DNA at the in vivo sites of integration. *J. Virol.* **63**:5319–5327.
 21. **Khan, E., J. P. G. Mack, R. A. Katz, J. Kulkosky, and A. M. Skalka.** 1990. Retroviral integrase domains: DNA binding and the recognition of LTR sequences. *Nucleic Acids Res.* **19**:851–860.
 22. **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.
 23. **Leavitt, A. D., R. B. Rose, and H. E. Varmus.** 1992. Both substrate and target oligonucleotide sequences affect in vitro integration mediated by human immunodeficiency virus type 1 integrase protein produced in *Saccharomyces cerevisiae*. *J. Virol.* **66**:2359–2368.
 24. **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.
 25. **Pecoraro, V. L., J. D. Hermes, and W. W. Cleland.** 1984. Stability constants of Mg²⁺ and Cd²⁺ complexes of adenine nucleotides and thionucleotides and rate constants for formation and dissociation of MgATP and MgADP. *Biochemistry* **23**:5262–5271.
 26. **Rowland, S.-J., and K. G. H. Dyke.** 1990. Tn552, a novel transposable element from *Staphylococcus aureus*. *Mol. Microbiol.* **4**:961–975.
 27. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
 28. **Sherman, P. A., and J. A. Fyfe.** 1990. Human immunodeficiency virus integration protein expressed in *Escherichia coli* possesses selective DNA cleaving activity. *Proc. Natl. Acad. Sci. USA* **87**:5119–5123.
 29. **Smith, J. S., and M. J. Roth.** 1993. Purification and characterization of an active human immunodeficiency virus type 1 RNase H domain. *J. Virol.* **67**:4037–4049.
 30. **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.
 31. **van Gent, D. C., C. Vink, A. A. M. Oude Groeneger, and R. H. A. Plasterk.** 1993. Complementation between HIV integrase proteins mutated in different domains. *EMBO J.* **12**:3261–3267.
 32. **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.
 33. **Vink, C., 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.
 34. **Vink, C., D. C. van Gent, Y. Elgersma, and R. H. A. Plasterk.** 1991. Human immunodeficiency virus integrase protein requires a sub-terminal position of its viral DNA recognition sequence for efficient cleavage. *J. Virol.* **65**:4636–4644.
 35. **Vora, A. C., M. L. Fitzgerald, and D. P. Grandgenett.** 1990. Removal of 3'-OH-terminal nucleotides from blunt-ended long terminal repeat termini by the avian retrovirus integration protein. *J. Virol.* **64**:5656–5659.