Characterization of Human Immunodeficiency Virus Type 1 Integrase Expressed in *Escherichia coli* and Analysis of Variants with Amino-Terminal Mutations

KAREN A. VINCENT,^{1†} VIOLA ELLISON,² SAMSON A. CHOW,² AND PATRICK O. BROWN^{1,2,3*}

Howard Hughes Medical Institute¹ and Departments of Pediatrics² and Biochemistry,³
Stanford University School of Medicine, Stanford, California 94305

Received 16 June 1992/Accepted 22 September 1992

Replication of a retroviral genome depends upon integration of the viral DNA into a chromosome of the host cell. The integration reaction is mediated by integrase, a viral enzyme. Human immunodeficiency virus type 1 integrase was expressed in *Escherichia coli* and purified to near homogeneity. Optimum conditions for the integration and 3'-end-processing activities of integrase were characterized by using an in vitro assay with short, double-stranded oligonucleotide substrates. Mutants containing amino acid substitutions within the HHCC region, defined by phylogenetically conserved pairs of histidine and cysteine residues near the N terminus, were constructed and characterized by using three assays: 3'-end processing, integration, and the reverse of the integration reaction (or disintegration). Mutations in the conserved histidine and cysteine residues abolished both integration and processing activities. Weak activity in both assays was retained by two other mutants containing substitutions for less highly conserved amino acids in this region. All mutants retained activity in the disintegration assay, implying that the active site for DNA cleavage-ligation is not located in this domain and that the HHCC region is not the sole DNA-binding domain in the protein. However, the preferential impairment of processing and integration rather than disintegration by mutations in the HHCC region is consistent with a role for this domain in recognizing features of the viral DNA. This hypothesis is supported by the results of disintegration assays performed with altered substrates. The results support a model involving separate viral and target DNA-binding sites on integrase.

An early event in the life cycle of all retroviruses, including the human immunodeficiency virus (HIV; 15, 50), is the integration of a double-stranded DNA copy of the viral genome into a chromosome of the host, a step that is necessary for efficient viral replication (for reviews, see references 6, 27, and 55). Genetic analyses have shown that the integrase-coding region, located at the 3' end of the *pol* gene (20, 41, 47), and sequences at the viral DNA termini (16, 17) are both required for integration of completely synthesized viral DNA.

In the cytoplasm of an infected cell, viral DNA and proteins are found within a subviral nucleoprotein complex (5, 7, 25). Isolated complexes can accomplish integration of viral DNA into an exogenous target in vitro (5, 7, 21, 24, 25). Linear viral DNA contained within the complexes is the direct precursor of the integrated provirus (7, 8, 26).

Analysis of the integration reaction in vitro has demonstrated that integrase is the only protein required for integration (12, 18, 31, 49). The integration reaction, as defined by in vitro studies, proceeds in three steps. The process is initiated by integrase-dependent cleavage of two nucleotides from the 3' end of each strand of the viral DNA (8, 32, 43). The next step (also catalyzed by integrase) is a concerted cleavage-and-ligation reaction during which a staggered cut is made in the target DNA and the recessed 3' ends of the viral DNA are linked to 5' ends of the target DNA at the cleavage site (8, 26). The resulting gaps in the target DNA are then repaired, perhaps by host enzymes, to create short direct repeats flanking the provirus, a hallmark of the retro-

viral integration reaction. Both of the integrase-catalyzed reactions proceed by a one-step transesterification mechanism (22). The cleavage and ligation reactions are reversible in vitro (14), providing evidence that integrase is indeed an enzyme.

To provide a source of integrase for our analysis of the HIV type 1 (HIV-1) integration reaction in vitro, we expressed the protein in *Escherichia coli* and purified it to near homogeneity by a procedure similar to those previously described (22, 35, 49, 58). This purified protein was used to carry out a detailed analysis of conditions that influence the 3'-end processing and cleavage-ligation reactions. Our results point out differences between the activity in vitro of HIV-1 integrase derived from a heterologous expression system acting on synthetic oligonucleotide substrates and authentic viral integrase acting as a component of the integrative complex from infected cells (21, 24).

We also initiated an effort to identify and characterize functional domains of HIV-1 integrase. Described here are five mutants with alterations within a phylogenetically conserved region located near the amino terminus of the protein. The landmarks of this region are pairs of histidine (His-12, His-16) and cysteine (Cys-40, Cys-43) residues that are highly conserved among retroviral and retrotransposon integrases. These have been demonstrated to coordinate a metal ion and form a defined structural domain (10), similar to the zinc finger commonly found in DNA-binding proteins (for reviews, see references 4, 23, and 30). Mutagenesis of selected amino acids within this region was carried out to assess their role in the various activities of integrase, including 3'-end processing, integration, and the reverse reaction (disintegration). In addition, these experiments address the hypothesis that these amino acids form part of a functionally

* Corresponding author.

† Present address: Genzyme Corporation, Framingham, MA 01701.

distinct structure that, by analogy with the zinc finger, might be involved in DNA binding. The data provide evidence for a model in which there are two separate DNA-binding sites on integrase.

MATERIALS AND METHODS

Cloning and expression of the HIV-1 integrase gene. HIV-1 integrase was expressed in *E. coli* by using the T7 RNA polymerase-promoter system (52). The integrase gene was derived from a plasmid clone of isolate HIVNL4-3 (1) by the polymerase chain reaction (44) and inserted into expression vector pT7-7. Oligonucleotide primers (Operon) for the polymerase chain reaction were prepared on an Applied Biosystems DNA synthesizer (PAN facility, Beckman Center, Stanford University, Stanford, Calif.). The primer used to derive the N terminus (5' GGAATCAGGAAACATATGTTTTAGATGGA 3') contains an *Nde*I site (underlined) for cloning; this site also includes an ATG initiation codon. The C-terminal primer (5' TGGTGTCTTTACTAAGCTTTTCCATGTGTTA 3') contains a *Hind*III site (underlined). Following confirmation of the primary structure of the clone by DNA sequence analysis as described previously (57), the construct was transformed into *E. coli* BL21 (DE3). Restriction enzymes and T4 DNA ligase were obtained from New England BioLabs, T4 polynucleotide kinase was from Pharmacia, and *Taq* polymerase was from Perkin-Elmer. All enzymes were used in accordance with manufacturer specifications. Standard cloning procedures were followed (45).

Purification of HIV-1 integrase. Integrase was purified in accordance with a modification of the procedure described by Sherman and Fyfe (49). A 10-liter volume of cells was grown at 30°C in LB containing 50 µg of ampicillin per ml. At an optical density at 600 nm of 1.0, IPTG (isopropyl-β-D-thiogalactopyranoside) was added to 0.25 mM and the cultures were incubated at 30°C for 3 h. Following harvesting, the cell pellet was frozen at -20°C. To begin enzyme preparation, the frozen bacterial pellet was suspended in 350 ml of lysis buffer (20 mM *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid [HEPES, pH 7.5], 1 mM EDTA, 5 mM dithiothreitol [DTT]). The cells were lysed by passing the suspension twice through a French press and then centrifuged (12,000 × *g*, 30 min). The insoluble pellet containing integrase was suspended in 100 ml of a buffer containing 100 mM NaCl in buffer A (20 mM HEPES [pH 7.5], 1 mM EDTA, 1 mM DTT); the suspension was Dounce homogenized 30 times and centrifuged (12,000 × *g*, 30 min). The washed pellet was suspended in 100 ml of 1 M NaCl in buffer B [20 mM HEPES [pH 7.5], 1 mM EDTA, 1 mM DTT, 10 mM 3-[(3-cholamidopropyl)-dimethyl-ammonio]-1-propanesulfonate [CHAPS]] and Dounce homogenized 50 times; the extract was then stirred slowly on ice for 1 h. Soluble integrase was recovered in the supernatant following centrifugation for 1 h at 85,500 × *g*. The supernatant was prepared for butyl Sepharose (Pharmacia) column chromatography by dropwise addition of 0.5 volume of a 2.4 M (NH₄)₂SO₄ solution [2.4 M (NH₄)₂SO₄ and 1 M NaCl in buffer A]. The sample was then loaded onto a column (8.2 ml) equilibrated in loading buffer [buffer A plus 0.8 M (NH₄)₂SO₄-1 M NaCl]. Following a wash in loading buffer (90 ml), bound integrase was eluted with a linear gradient (82 ml) from the loading buffer to buffer B plus 100 mM NaCl. Fractions containing integrase were pooled and diluted ninefold in buffer B. This sample was then loaded onto tandem DEAE Sepharose (8.5 ml) and heparin Sepharose (4.6 ml) columns (Pharmacia)

equilibrated in loading buffer (buffer B plus 100 mM NaCl). Under these conditions, integrase first flowed through DEAE and then bound to heparin Sepharose. The heparin column was washed (90 ml) in loading buffer, and bound integrase was eluted with a linear gradient (46 ml) from 100 mM to 1 M NaCl in buffer B. The integrase peak eluted at approximately 600 mM NaCl. In the final step, pooled fractions were subjected to chromatography over a Sephacryl S-100 (Pharmacia) column (120 ml) equilibrated in buffer B plus 300 mM KCl.

The final concentration of integrase was determined by amino acid analysis. The results of this analysis agreed with the predicted amino acid composition of integrase. Purified integrase was stored at -70°C in S-100 buffer.

Molecular weight determination. The molecular weight of HIV-1 integrase purified from *E. coli* was estimated by both gel filtration chromatography and glycerol gradient sedimentation. A Sephacryl S-200 column (48 by 0.9 cm; 40-ml column volume) was equilibrated with buffer X (20 mM HEPES [pH 7.5], 1 mM DTT, 1 mM EDTA, 0.05% Nonidet P-40, 600 mM NaCl) and run at a flow rate of 4.4 ml/h. Protein standards (thyroglobulin, bovine gamma globulin, chicken ovalbumin, equine myoglobin, vitamin B-12; Bio-Rad) and integrase (69 µg) were analyzed in separate runs. Elution volumes were determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and densitometry (Molecular Dynamics).

Samples for glycerol gradient sedimentation were either 30 µg of integrase alone, 30 µg each of integrase, carbonic anhydrase, and bovine serum albumin (BSA), or 30 µg each of carbonic anhydrase, BSA, and chicken ovalbumin; the sample volume was 100 µl. The samples were loaded at 4°C onto 5 to 20% glycerol gradients (4.2 ml) in buffer X. Centrifugation was done at 4°C for 16 h at 485,000 × *g* in an SW 60Ti rotor (Beckman). Fractions were collected from the top of the gradient and analyzed by SDS-PAGE and densitometry.

Assays for integrase activity. The substrate used to assay the integration and end-processing activities of the HIV-1 integrase was a double-stranded oligonucleotide containing sequences derived from the long terminal repeat terminus (U5 end). Oligonucleotides to be used as substrates were purified by electrophoresis through a 20% polyacrylamide gel before use. The substrate was prepared by labelling the 5' end of the processed strand (oligonucleotide U5C2; 5' AATTCGATACTATGTGGAAAATCTCTAGCAGT 3') and annealing it with an excess of its complement (oligonucleotide U5C1; 3' GCTATGATACACCTTTTAGAGATCGTCA 5'). This substrate has four unpaired bases at the 5' end of the processed strand. Control experiments showed that the presence of the excess complementary oligonucleotide had no effect on reaction efficiency. T4 polynucleotide kinase (Pharmacia) was used to label the 5' end with [γ -³²P]ATP (3,000 Ci/mmol; ICN); the oligonucleotide was typically labelled to a specific activity of approximately 10⁶ cpm/pmol. In the integration assay, 0.1 pmol of the substrate was incubated with 2 pmol of integrase in reaction buffer (20 mM HEPES [pH 7.5], 1 mM DTT, 0.05% Nonidet P-40, 5 to 20 mM NaCl, 15 mM MnCl₂) for 1 h at 30°C. The total reaction volume was 20 µl. Reaction products were visualized by electrophoresis through a 20% polyacrylamide-8 M urea gel and autoradiography. An assay for integration activity alone was performed with a substrate containing a prerecessed 3' end on the processed strand (oligonucleotide U5C2-2; 5' AATTCGATACTATGTGGAAAATCTCTAGCA 3'); this was annealed with oligonucleotide U5C1.

The disintegration assay was performed as previously described (14). The following oligonucleotide was used to create the "dumbbell" substrate: 5' TGCTAGTTCTAG CAGGCCCTTGGGCCGGCGCTTGCGCC 3'. Quantitation of integrase activity in both integration and disintegration assays was carried out by densitometry or by phosphorimager scanning (Molecular Dynamics).

Site-directed mutagenesis. Point mutations were introduced into the integrase gene in accordance with the polymerase chain reaction method described by Hemsley et al. (28). Constructs were first transformed into *E. coli* HB101 to provide plasmid DNA for restriction and sequence analysis. Each mutant gene was sequenced completely to confirm the desired mutation and the absence of secondary mutations. The mutant clones were then transformed into *E. coli* BL21(DE3).

Mutant integrase proteins were purified by a modification of the procedure described above including only the butyl and heparin Sepharose column chromatography steps. Mutant integrases were purified to 75 to 85% homogeneity by this protocol. Activities of the mutant proteins were compared with those of a wild-type enzyme purified by the same procedure.

RESULTS

Purification of HIV-1 integrase expressed in *E. coli*. HIV-1 integrase expressed in *E. coli* by using the T7 expression system is identical to authentic viral integrase, except for a methionine residue that precedes the phenylalanine normally found at the N terminus of the viral protein. Low-speed centrifugation ($12,000 \times g$) of the bacterial lysate yielded an insoluble pellet containing integrase. Approximately 70% of the integrase protein was dissolved by extraction of the pellet with 1 M NaCl and 10 mM CHAPS. Integrase was purified by a procedure involving butyl, DEAE, and heparin Sepharose chromatography steps. The final step was gel filtration through a Sephacryl S-100 column (Fig. 1). At the final stage, the full-length integrase protein made up approximately 80% of the total protein. As shown in Fig. 1B, the remainder of the contaminants are slightly faster migrating polypeptides. These are truncated derivatives of integrase, as they reacted with integrase-specific antisera on a Western immunoblot (data not shown); however, the exact structure of these polypeptides is unknown. Integrase expressed in *E. coli* comigrated with authentic viral integrase on SDS-PAGE with an apparent molecular mass of 32 kDa.

Assays of integrase activity, including processing, integration, and disintegration, were performed on fractions from each of the columns (except DEAE Sepharose). These assays showed that all activities coeluted with the peak of integrase protein (Fig. 1 and data not shown). Nonspecific nuclease activity was apparent in these assays in early steps of purification; it was removed by successive column steps. Gel filtration allowed resolution of the integrase peak from a contaminating activity that generated a -1 product in the 3'-end processing assay (Fig. 1D). However, a -2 band was not detected in the fractions containing the nuclease (even in a longer exposure of the autoradiogram). This is of particular significance for the analysis of mutant integrase preparations, which were not subjected to S-100 chromatography. Integrase chromatographed similarly relative to calibration standards on the preparative S-100 and analytical S-200 columns (see below).

Table 1 summarizes the purification. Integrase activity was quantitated at each step by the disintegration assay. A

unit of activity was defined as the amount of enzyme required to convert 2.5% of the substrate (0.1 pmol per reaction) to the product in 15 min under standard reaction conditions (Table 1). This level of activity is within the linear range of the disintegration assay (data not shown). Total activity values in the early steps are minimal estimates because contaminating nucleases in these samples degrade the reaction products. For this reason, total activity increased through the butyl Sepharose step and the yield in later steps was calculated with respect to the total activity in the butyl Sepharose pool. Specific activity appeared to decrease with each subsequent chromatography step; this result may indicate that integrase activity is not stable at 4°C under the conditions utilized. Nevertheless, these columns were necessary to remove residual contaminating nuclease activity that persisted up to the Sephacryl S-100 step (Fig. 1D).

The molecular mass of HIV integrase determined by SDS-PAGE is approximately 32 kDa, which is consistent with that predicted by the amino acid sequence. The Stokes radius determined by gel filtration through Sephacryl S-200 (Fig. 2A) was 4.37 nm, and the sedimentation coefficient ($s_{20,w}$) estimated by glycerol gradient sedimentation (Fig. 2B) was 3.13×10^{-13} . The molecular mass and frictional coefficient (f/f_0) calculated from these parameters were 57,370 daltons and 1.7, respectively (50), assuming a partial specific volume of 0.725 ml/g. We cannot rule out the possibility that the shorter polypeptides present in the preparation also participate in protein-protein interactions with full-length integrase and have some effect on the calculated values. However, these data suggest that under the conditions employed, integrase behaves as a markedly nonspherical dimer.

Effects of various reaction conditions on the integration and processing activities of HIV-1 integrase. Integrase activity was analyzed in an in vitro assay using double-stranded oligonucleotide substrates consisting of sequences derived from the long terminal repeat termini (18, 32, 49). These serve as both the substrate and the target in the reaction. In most assays, the substrate was a 32-mer containing sequences from the U5 end (see Materials and Methods). As others have noted (11, 35, 36), the integration reaction proceeded less efficiently with a substrate composed of sequences from the U3 end (data not shown). In the first step of the integration reaction (referred to as 3'-end processing), two nucleotides are removed from the 3' end of the substrate. Cleavage occurs immediately 3' to a conserved CA dinucleotide, yielding a 30-nucleotide product.

Figure 3A shows the effects of various salt concentrations, pHs, temperatures, and incubation times on the in vitro integration and 3'-end processing activities of HIV-1 integrase. In these experiments, integration activity was assayed independently of processing by using a preprocessed substrate. Incubation times were kept short (15 min) to examine the effects of various conditions on the initial rate rather than the extent of the reaction. NaCl had an inhibitory effect on both activities; a significant reduction in both activities was observed at 50 mM NaCl. Both integration and processing activities peaked at pH 7.8, though integration exhibited a broader pH curve. Both reactions occurred over a broad range of temperatures. Maximum processing and integration activities were observed at 42°C. During a time course analysis, the rate of integration began to plateau after 30 min and processing activity plateaued after 1 h at 30°C. Both integration and processing activities increased linearly with enzyme concentration up to 50 nM (data not shown).

FIG. 1. (A) Elution profile of the Sephacryl S-100 gel filtration column. Shown are protein concentration, disintegration activity, and processing activity plotted against elution volume. The protein concentration in each fraction was estimated by using the Bio-Rad protein assay reagent. Units of activity were calculated from the disintegration assay as described in Results. Processing activity is presented as percent conversion of the substrate into the product as quantitated by densitometry. (B) Coomassie-stained SDS-PAGE gel of fractions containing protein. Lanes are labelled with the fraction number and correspond to datum points in panel A. Labels at the left correspond to the positions of molecular weight standards run on the same gel. Fraction 12 was aliquotted and frozen at -70°C as a stock of purified wild-type protein. (C) Results of a disintegration assay performed with each fraction. Disintegration reactions were carried out as previously described (14). The products were electrophoresed in a denaturing 15% polyacrylamide gel. The filled arrow indicates the position of the 5'-end-labelled strand of the substrate, which migrated as a 16-nucleotide oligomer on the denaturing gel. The open arrow corresponds to the 30-nucleotide-long product. (D) Results of a 3'-end processing assay performed with each fraction. The reaction products were electrophoresed on a denaturing 20% polyacrylamide gel. The filled arrow corresponds to the 5'-end-labelled input substrate. The stippled and open arrows indicate the positions of the -1 and -2 products, respectively. The latter is the product of the integrase-dependent 3'-end processing reaction.

TABLE 1. Purification of HIV-1 integrase from *E. coli*

Step	Total Protein (mg)	Total activity (U, 10^3) ^a	Sp act (U/mg, 10^3) ^a	Yield (%)
Lysate	6,600	ND	ND	
12,000 $\times$ g pellet	170	130	0.8	
Wash pellet	130	324	2.5	
NaCl extract	46.3	224	4.8	
Butyl pool	15.9	567	36	100
Heparin pool	3.3	89	27	16
S-100 pool	2.1	46	22	8

^a Activity was quantitated in the disintegration assay; autoradiograms were scanned by densitometry. A unit is defined as the amount of enzyme required to convert 2.5% of the substrate into the product in 15 min at 37°C (reaction conditions: 20 mM HEPES [pH 7.5], 1 mM DTT, 0.05% Nonidet P-40, 10 mM MnCl_2 , 27 mM NaCl). This level of activity is within the linear range of the assay. ND, not done.

Divalent cations were absolutely required for integrase activity. Both integration and processing occurred in the presence of Mn^{2+} ; optimal activity was observed with 25 mM MnCl_2 (Fig. 4B). In contrast, while processing activity was readily detectable when the divalent cation was Mg^{2+} , only trace levels of integration products were observed with long exposure times. Furthermore, no activity was detected when other divalent cations (Co^{2+} , Ca^{2+} , and Zn^{2+}), spermine, or spermidine were tested in the reaction as alternatives to Mn^{2+} (data not shown).

The activity of integrase *in vitro* was not affected when the reaction mixture was supplemented with either BSA (100 mg/ml), polyethylene glycol (5%), glycerol (10%), or ATP (1 mM) (data not shown).

Analysis of amino-terminal mutants of HIV-1 integrase. Amino acid substitutions were introduced within the conserved HHCC region at the N terminus of the protein, which bears some similarity to the zinc finger motif found in many DNA-binding proteins (10, 29, 33). Figure 4A shows the

FIG. 2. (A) Determination of the Stokes radius of HIV-1 integrase. Gel filtration data were plotted as described by Siegel and Monty (50). The Stokes radius standards used to calibrate the column were carbonic anhydrase (●), ovalbumin (○), BSA (■), and bovine gamma globulin (□). The Stokes radius of integrase relative to the standards is as indicated by the arrowhead. (B) Determination of the sedimentation coefficient of HIV-1 integrase. Integrase was sedimented through a 5 to 20% glycerol gradient as described in Materials and Methods. The standards used to calibrate the column were carbonic anhydrase (●), ovalbumin (○), and BSA (■). The position of the integrase peak within the gradient is as indicated by the arrowhead.

amino acid changes made in each mutant. Two mutants, IN(H12A, E13L) and IN(S39T, C40A), include substitutions in the invariant histidine and cysteine residues that define this region of the integrase protein. IN(P29A) substituted alanine for a proline residue that is conserved in lentivirus integrases and predicted to form a turn or chain reversal within proposed zinc finger structures (10, 29). Three adjacent amino acids were substituted in IN(V32G, A33L, K34A); these included Ala-33, which is conserved at this position in many retroviral integrases (33). The final mutant, IN(G47L, E48A), differs from the wild type in two relatively unconserved amino acids just C terminal to the HHCC region. Because most of the mutants contain multiple substitutions, the phenotypes observed cannot strictly be attributed to a single amino acid change; however, in mutants IN(H12A, E13L) and IN(S39T, C40A), mutation of the phylogenetically invariant amino acid (His-12 and Cys-40, respectively) is presumably more significant. This assumption is supported by the finding that a single amino acid substitution at Ser-39 (S39A) had no effect on processing,

integration, or disintegration activity (54). All mutants except IN(G47L, E48A) were less soluble than the wild type in 1 M NaCl, but all bound to butyl and heparin Sepharose in a manner similar to that of wild-type integrase, suggesting that the mutations did not have a radical effect on the conformation of the protein.

The activities of each mutant were examined in three assays: 3'-end processing, integration (using a prerecessed substrate), and disintegration (14). Table 2 summarizes the results. Similar results were obtained with two independently purified preparations of each mutant. Mutant proteins carrying substitutions in the conserved histidine and cysteine residues lacked both 3'-end processing and integration activities. Weak processing activity was observed with IN(P29A) and IN(V32G, A33L, K34A); a low level of integration activity was detected with mutant IN(P29A) but not with IN(V32G, A33L, K34A). Mutant IN(G47L, E48A) retained activity similar to that of the wild-type protein in both assays. In contrast, all of the mutants retained wild-type levels of activity in the disintegration assay (Fig. 5). This result implies that none of the mutations affects residues directly involved in the cleavage and ligation activities of integrase. The minor bands visible between the substrate and product in panels A and B and migrating more slowly than the substrate in panel B are likely to be the products of reintegration of the liberated viral end following disintegration. The presence of these products in reactions catalyzed by the mutant enzymes indicates that they are able to carry out integration under these conditions, although no integration activity was detected with the standard integration assay. The possible implications of this result for understanding of enzyme-DNA complex formation are under investigation.

Why is disintegration activity insensitive to mutation in the HHCC region? Experiments were done to investigate why some of the mutants were unable to carry out integration efficiently but still retained the ability to catalyze the reaction in reverse. One possible explanation for this result is that mutations in the HHCC region impair interactions between integrase and viral DNA sequences. This might be expected to interfere with activity in the integration assay, in which the substrate is composed entirely of viral DNA sequences. However, the protein might still bind the disintegration substrate via interaction with the target or non-specific DNA component of the substrate.

To ascertain further the role of the HHCC region, additional disintegration assays were conducted, using substrates that had altered structures. Figure 5 shows an experiment comparing each protein's disintegration activity with that of the standard substrate (panels A and B) and that of a dumbbell substrate (panel C). The latter substrate differs from the Y substrate in that it is much smaller (5 bp of the viral sequence and 10 bp of the target versus 19 and 30 bp, respectively in the Y substrate) and is composed of a single oligonucleotide rather than four. Wild-type integrase activity is greater with the dumbbell substrate than with the Y substrate. Under optimal conditions, >80% of the dumbbell is converted into the product, as opposed to 40% with the Y substrate (13, 14). Unlike the Y substrate, the dumbbell is disintegrated at a reduced level by integrases with mutations in the HHCC region. The weakest activity was observed with IN(S39T, C40A), while mutation outside of the proposed structural domain (G47L, E48A) caused little or no reduction in disintegration activity. Moderate activity with the dumbbell substrate was retained by mutants IN(H12A,

FIG. 3. Influence of reaction conditions on processing and integration activities of purified HIV-1 integrase *in vitro*. Processing activity was assayed by using as the substrate a double-stranded oligonucleotide composed of U5C2 annealed to U5C1; integration activity was assayed by using a preprocessed substrate (U5C2-2 annealed to U5C1). (A) Effects of NaCl concentration, pH, temperature, and time on the efficiency of the two reactions. Reactions were carried out as described in Materials and Methods, with the variables indicated. Activity was quantitated by densitometry and is presented as percent conversion of the substrate into the product. (B) Activity of the enzyme in reactions containing Mn^{2+} or Mg^{2+} as the divalent cation. In the latter graph, only processing is shown; very weak integration activity was observed when the preprocessed substrate was used.

FIG. 4. Mutations introduced within the HHCC region of HIV-1 integrase. (A) Amino acid changes made in each mutant. The integrase coding region (289 amino acids) is represented by the solid bar. Highlighted above the bar are the conserved pairs of histidine and cysteine residues that define this region of the protein. (B) Schematic of the HHCC region, depicted as a zinc finger domain. Amino acids substituted in the mutants are circled.

E13L), IN(V32G, A33L, K34A), and IN(P29A); the former two mutants were approximately half as active as the latter.

Disintegration of both the dumbbell and the Y substrates by mutant proteins IN(H12A, E13L), IN(P29A), IN(V32G, A33L, K34A), and IN(S39T, C40A) was hypersensitive to the salt concentration in the reaction mixture compared with the same reaction catalyzed by the wild-type protein or mutant IN(G47L, E48A). Disintegration activity with the Y substrate decreased two- to sevenfold when the NaCl concentration was increased from 20 to 50 mM in the reaction (compare Fig. 5A with 6A). A similar increase in salt concentration had an even more dramatic effect on disintegration activity with the dumbbell substrate, rendering the mutant proteins unable to use this structure as a substrate for disintegration (compare Fig. 5C with 6B).

The sensitivity of disintegration activity with the dumbbell substrate to mutation in the HHCC region might be explained if these mutations interfere with binding to the viral portion of the substrate while compensatory binding to the target region (as proposed for the Y substrate) is prevented either because this region is too small in the dumbbell (10 bp versus 30 bp in the Y substrate) or because binding to this

substrate is dependent upon interaction with both the viral and target regions. Two hybrid disintegration substrates were designed to address this possibility. In one, db-V (Fig. 6C and D), the viral portion of the substrate has the short dumbbell configuration but the target is the 30-bp region derived from the Y substrate; the second, db-T (Fig. 6E), has the reciprocal structure in which the target portion has the short dumbbell configuration but the viral strands are full length. Reactions shown in the figure were performed at a high salt concentration (50 mM NaCl) to accentuate the defective activity of the mutants; qualitatively similar results were obtained with a low salt concentration (20 mM NaCl). Reinstating the full-length target strands, as in substrate db-V (Fig. 6C and D), restored activity to the level observed with the Y substrate. In contrast, only a low level of activity in all of the mutants was observed with the db-T substrate (Fig. 6E).

Additional assays were performed to determine whether a particular region of the viral DNA is recognized by the HHCC domain. Figure 7B shows the results of an experiment performed with a substrate in which the conserved CA dinucleotide in the viral sequence was substituted with TC. The wild-type enzyme carried out disintegration with this substrate, but at a level reduced sixfold relative to the standard substrate (14). The CA-to-TC substitution also impaired disintegration by the mutant proteins, implying that the mutants retained the ability to interact specifically with the CA dinucleotide. Furthermore, the mutant enzymes appeared to be hypersensitive to this alteration in the substrate sequence. Disintegration of the modified substrate was reduced 16- to 120-fold, depending on the mutant, relative to the standard substrate.

Disintegration assays were also carried out with a substrate from which virtually all viral DNA sequences were removed, leaving only a single adenine nucleotide (Fig. 6C). Previous analyses (14) have shown that this structure is used as a substrate for disintegration, but only about 10% as efficiently as the standard Y substrate. If, as suggested by the previous results, the HHCC region is only involved in recognizing viral DNA end sequences internal to the CA

TABLE 2. In vitro activities^a of HHCC mutants

Integrase	Processing	Integration	Disintegration ^b
IN(wt)	++++	++++	++++
IN(H12A, E13L)	—	— ^c	++++
IN(P29A)	+	+ ^c	++++
IN(V32G, A33L, K34A)	+	— ^c	++++
IN(S39T, C40A)	—	— ^c	++++
IN(G47L, E48A)	++++	++++	++++

^a Relative activities were quantitated from autoradiograms by densitometry. +, 0 to 10% of wild-type activity; —, no activity; +++++, wild-type level of activity.

^b Disintegration activity as assayed with the Y substrate under low salt concentration conditions (20 mM NaCl).

^c Although little or no integration activity was observed for the mutant in the integration assay, a ladder-like array of bands appeared between the substrate and product in the disintegration assay, indicating that these mutants can carry out reintegration following disintegration (see text).

FIG. 5. Disintegration assays carried out with the standard Y substrate and the dumbbell substrate. Reactions were performed at a low salt concentration (20 mM NaCl). The structure of the substrate used in each assay is shown below the panel; bold lines represent viral DNA sequences, and the filled circle represents a labelled 5' end. The open arrowheads indicate the position of the substrate, and the filled arrowheads designate the position of the disintegration product for each set of reactions. Reactions were performed with the standard Y substrate (14) labelled on the 16-mer (A), yielding a 30-nucleotide product, and on the 33-mer (B), yielding a product 19 nucleotides long. Panel C shows reactions carried out with the dumbbell substrate. The product in these reactions is 14 nucleotides long. Bands appearing between the substrate and product in panels A and B and also those that migrate more slowly than the substrate in panel B are believed to arise from reintegration of a liberated viral end following disintegration (see text). Lanes labeled -IN and WT in all panels are reactions performed without integrase and with the wild-type enzyme, respectively. Lanes 1, IN(H12A, E13L); 2, IN(P29A); 3, IN(V32G, A33L, K34A); 4, IN(S39T, C40A); 5, IN(G47L, E48A).

dinucleotide, then disintegration activity with this substrate should be equivalent for all of the mutants and similar to that of the wild type. As shown in Fig. 7C, disintegration activity was relatively homogeneous (percent conversion of the substrate ranged from 0.3 to 1.7%). Similar results were obtained with a substrate in which the A was replaced with a C, which should remove the remaining influence of a specific viral DNA sequence on disintegration activity (Fig. 7D).

An additional experiment was done to address the question of which feature(s) of the viral DNA might be recognized by the HHCC domain. Shown in Fig. 8 are reactions performed with substrates in which only the conserved CA dinucleotide was single stranded (panel B), the entire viral DNA portion was single stranded (panel C), or the viral DNA was single stranded and reduced in length to 5 bases (panel D). The disintegration activity of the mutant proteins was reduced when the CA dinucleotide was made single stranded (panel A), corroborating the result obtained when the CA dinucleotide was substituted (Fig. 7B). As observed in the previous experiment, the mutants were hypersensitive to this alteration; activity of the mutant proteins was reduced 10- to 25-fold, whereas wild-type activity decreased only 3-fold. Making the entire viral DNA portion single-stranded further reduced the disintegration activity (approximately twofold) observed with the wild type and mutants IN(G47L,

E48A) and IN(P29A), both of which retained at least some integration activity. In contrast, the disintegration activity of the remaining mutants was unaffected, suggesting that these mutants were unable to recognize this further alteration in the viral DNA structure internal to the CA dinucleotide. When the length of the viral strand was further reduced to 5 bases, a result similar to that shown in Fig. 7C and D was obtained; disintegration activities among the mutant and wild-type integrases became more homogeneous. We do not have an explanation for why mutant IN(S39T, C40A) was reproducibly hypersensitive to shortening of the viral strand (compare also Fig. 5A and C and 7A and C). Perhaps this most severe alteration in the HHCC domain abolished a residual interaction between the other HHCC mutants and some component of this 5-bp single-stranded viral DNA. A mutant in which the HHCC region was deleted (amino acids 13 to 42) had properties identical to those of mutant IN(S39T, C40A) in all experiments with the various altered substrates (56).

DISCUSSION

We expressed HIV-1 integrase in *E. coli* and characterized the influence of reaction conditions on both the 3'-end processing and cleavage-ligation activities of the purified protein. Reaction requirements for the bacterially derived enzyme differ from those reported for viral integrase. Of

FIG. 6. Disintegration assays carried out with the dumbbell and hybrid dumbbell substrates. Reactions were performed in a high salt concentration (50 mM NaCl). The structure of the substrate used in each assay is shown below the panel; bold lines represent viral DNA sequences, and the filled circle represents a labelled 5' end. (A) Reactions performed with the standard Y substrate (14). (B) Reactions carried out with the dumbbell substrate. (C and D) Results obtained with a hybrid substrate (db-V) in which the viral portion is in the dumbbell configuration and the target portion is derived from the Y substrate. In panel C, the 5' end of the 16-mer is labelled with ^{32}P , while in panel D, the label appears on the hybrid virus-target strand. (E) Reactions carried out with the reciprocal structure (db-T), in which the viral portion is derived from the Y substrate and the target component is in the dumbbell configuration. The open arrowheads indicate the position of the substrate, and the filled arrowheads designate the position of the disintegration product for each set of reactions. Bands migrating more slowly than the substrate in panel D (lanes WT and 5) and the ladder of bands visible between the substrate and product in panels A, C, and E are believed to arise from reintegration of the viral DNA strand following disintegration. Lanes labelled - and WT in all panels are reactions performed without integrase and with the wild-type enzyme, respectively. Lanes 1, IN(H12A, E13L); 2, IN(P29A); 3, IN(V32G, A33L, K34A); 4, IN(S39T, C40A); 5, IN(G47L, E48A).

particular note is the nearly absolute dependence of the purified bacterial enzyme on Mn^{2+} for efficient integration activity, while 3'-end processing can occur in the presence of either Mn^{2+} or Mg^{2+} . In contrast, Mg^{2+} fulfills the divalent cation requirement in assays that employ cytoplasmic extracts from infected cells as a source of HIV integrase (21, 24). Similar results have been obtained with murine leukemia virus integrase (7, 18, 26). Avian sarcoma-leukosis virus integrase purified from virus particles carries out integration more efficiently in the presence of Mn^{2+} (31), similar to the result obtained with overexpressed proteins. In light of this finding, a possible explanation for the altered divalent cation requirement of the bacterially derived enzyme might be that it is a distinctive property of the free enzyme, outside of a nucleoprotein complex.

Our physical characterization of the bacterially expressed enzyme provided evidence to suggest that under the experimental conditions employed, integrase behaves as a non-spherical dimer. Sherman and Fyfe (49) have also reported that integrase sediments as a dimer during glycerol gradient centrifugation. Kinetic and ultracentrifugation analyses have also suggested that the minimal functional form of integrase is a dimer (29a). In contrast, others have found that integrase behaves as a monomer during gel exclusion chromatography (36, 53). These conflicting results might be explained by the different conditions employed in each experiment, specifically, salt and detergent concentrations, both of which have profound effects on the behavior of integrase in these procedures. The experiments reported here are the first in which both glycerol gradient centrifugation and gel exclusion chromatography were performed under identical conditions.

Studies of HIV-1 integrase by site-directed mutagenesis may allow identification of specific regions and/or amino

FIG. 7. Comparison of mutant disintegration activities achieved with the standard and modified substrates. (A) Series of reactions performed with the wild-type substrate. (B) Assay carried out with a substrate in which the conserved CA dinucleotide was substituted with a TC dinucleotide. (C) Reactions carried out with a disintegration substrate retaining only a single A from the viral sequence. (D) Disintegration reactions performed with a similar substrate in which the A was substituted with a C. The structure of the substrate used in each assay is depicted below the panel. In all cases, the substrate (open arrowhead) was radiolabelled at the 5' end of the 16-mer and the products (filled arrowhead) were 30 nucleotides long. Lanes labelled - and WT in all panels are reactions performed without integrase and with the wild-type enzyme, respectively. Lanes: 1, IN(H12A, E13L); 2, IN(P29A); 3, IN(V32G, A33L, K34A); 4, IN(S39T, C40A); 5, IN(G47L, E48A). Note that the bands migrating between the substrate and product are not as apparent in reactions carried out with the substrate containing the CA-to-TC substitution, as might be expected if these bands resulted from reintegration following disintegration.

FIG. 8. Disintegration assays carried out with mutant proteins and substrates with modified structures (shown schematically below the panels). Reactions were performed with the standard Y substrate (A) and with substrates in which the conserved CA dinucleotide was single stranded along its entire length (B), single stranded and reduced in length to 5 bases (C), or single stranded and reduced in length to 5 bases (D). In all cases, the substrate (open arrowhead) was radiolabelled at the 5' end of the 16-mer and the products (filled arrowhead) were 30 nucleotides long. Lanes are labelled as in Fig. 7.

acids involved in the various activities of the enzyme. An initial analysis was designed to determine the functional significance of a region at the N terminus of the protein defined by pairs of histidine and cysteine residues (HHCC region). These are absolutely conserved within the integrase-coding regions of retroviruses and related elements, such as copia and Ty (33). It has been noted (29) that these amino acids might constitute part of a metal-binding domain analogous to the zinc finger found in many DNA-binding proteins (4, 23, 30, 34). Structural analyses have revealed that metal ions are bound with equimolar stoichiometry by a peptide composed of the first 55 amino acids of HIV-1 integrase (10). One conserved histidine residue (His-12) and the two cysteines (Cys-40 and Cys-43) were shown to participate in metal binding, which results in a transition to a more highly ordered structure (10).

The prototypic zinc finger, first identified in *Xenopus laevis* transcription factor TFIIIA (9, 38), is composed of the sequence Cys-Xaa₄-Cys-Xaa₁₂-His-Xaa₃-His. As noted previously (33), this differs from the corresponding region of integrase in three major respects. (i) The order of the conserved pairs of residues is HHCC in integrase but CCHH in TFIIIA. Although the order of the conserved pairs is different for the integrases, spacing between residues of each pair conforms to the consensus (42). Alternate zinc finger types have also been characterized. For example, the finger domains of the retroviral nucleocapsid and the glucocorticoid receptor, which have the forms CCHC (2) and CCCC (3), respectively, fold into structures completely different from that of the TFIIIA domain (30). (ii) the number of amino acids between the pairs of conserved residues is much larger for the integrase motif: 23 amino acids for HIV integrase versus 12 for the prototypic zinc finger. (iii) Functional finger domains almost always appear in multiple copies; for example, TFIIIA contains nine repeating units and an equivalent number of zinc atoms per molecule (23). In contrast, the HHCC region is the only such motif in the integrase-coding region. The nucleocapsids of some retroviruses also contain only one copy of the CCHC domain (2, 4).

Mutations in and surrounding the conserved His (His-12) and Cys (Cys-40) residues of HIV-1 integrase had the most drastic effect on activity, providing evidence for their par-

FIG. 9. Model for integrase activity. (A) Important functional domains in integrase. Integrase is proposed to have two DNA-binding sites, one for viral DNA and a separate one for target DNA. The protein has a single active site. (B) In the 3'-end processing reaction, viral DNA is bound within the viral DNA binding site and a phosphodiester bond is made susceptible to nucleophilic attack by water. (C) In the integration reaction, target DNA is bound and a phosphodiester bond is activated for nucleophilic attack by the free 3'-hydroxyl on the viral DNA end. This reaction is carried out by a dimer in which each subunit is responsible for attaching a single viral DNA end to target DNA. Although not depicted here, it is possible that both subunits contribute portions of each viral DNA-binding site.

ticipation in a functionally important structure. An essential role for the His and Cys residues is also consistent with the pH profile of integration activity. The importance of Cys-40, in particular, is suggested by the finding that a mutant containing a deletion of the entire HHCC region (amino acids 13 to 42) possessed a phenotype similar to that of mutant IN(S39T, C40A) in all experiments (56). Mutation of the less conserved amino acids between the two pairs of His and Cys residues (P29A and V32G, A33L, K34A) also reduced processing and integration activity, indicating that

these amino acids also contribute to the function of this region of the protein. The significance of this region *in vivo* has been demonstrated by a genetic analysis of the murine leukemia virus integrase-coding region (19). Mutants containing linker insertions within or adjacent to the conserved His and Cys pairs were unable to generate stable proviruses, although they retained the ability to synthesize viral DNA.

In a previous analysis of the function of this region in avian sarcoma-leukosis virus integrase (33), mutant integrases with conservative substitutions at each of the conserved histidines exhibited reduced processing (~50%) and integration activities relative to the wild type. Similarly, HIV-1 integrase mutants possessing both conservative and nonconservative substitutions at His-16 (H16C and H16V, respectively) retained nearly wild-type processing and integration activities (37). Mutational analysis of HIV-2 integrase has shown that substitution of His-12, Cys-40, and Cys-43 reduced processing activity more than it reduced integration activity (54); notably, a mutant containing a substitution of a serine for Cys-40 retained a nearly wild-type level of integration activity. The discrepancy between these results and the result obtained with mutants IN(H12A, E13L) and IN(S39T, C40A) may be attributable to the fact that the latter are double mutants or to the specific substitutions that were made in each case (or both). It is relevant to note that substitution of Ser-39 with an alanine had no effect on processing, integration, or disintegration activity (54). In addition, the salt sensitivity of the mutant proteins examined in this work emphasizes the dependence of the results on the specific reaction conditions employed.

The sensitivity of the mutant enzymes to defined alterations in the substrate suggests specific functions that the HHCC region might serve and rules out others. The retention of disintegration activity by all of the mutants, even those with radical substitutions for phylogenetically invariant His-12 and Cys-40, argues strongly against an essential role in catalysis *per se*. Moreover, since the activity of the mutant enzymes was strongly dependent on the invariant CA dinucleotide at the viral DNA end (Fig. 7 and 8), an intact HHCC motif is not essential for recognition of this substrate feature. A structural element similar to a zinc finger might be expected to function in DNA binding. However, all of the mutants characterized here retained DNA-binding activity, as reflected by their ability to promote disintegration. These results imply that the HHCC region cannot be the sole DNA-binding domain of the protein.

The working model for integrase activity shown in Fig. 9 forms a framework for further discussion of the mutant phenotypes. Important aspects of this model are that (i) integrase has two DNA-binding sites, one for viral DNA and one for target DNA; (ii) there is a single active site in the protein; and (iii) the integration reaction is carried out by a dimer of integrase molecules; each subunit is responsible for mediating the attachment of a single viral DNA end. The concerted integration of both viral DNA ends into precisely spaced sites in the target DNA implies that interactions between integrase monomers are likely to be important for this reaction. That such interactions may also be important for 3'-end processing activity is also suggested by the finding that mutations in one end of murine leukemia virus DNA that block its processing may also result in impaired processing of a completely wild-type opposite end (40).

In the 3'-end processing reaction (Fig. 9B), a viral DNA end is bound in the viral DNA-binding site so as to position the phosphodiester bond following the CA dinucleotide in the active site. The essential catalytic activity of the active

site is to make this bond susceptible to nucleophilic attack by water or an alcohol (22, 59). Similarly, in the integration reaction (part C), the free 3' hydroxyl at the viral DNA end acts as a nucleophile to attack a susceptible phosphodiester bond in a target DNA bound within the target DNA-binding site. The demonstration of a single-step transesterification mechanism for both the 3'-end processing and integration activities of integrase (22) and the very similar structural requirements for the DNA substrates provide strong circumstantial evidence for a single active site in the protein. Other models for the integrase-mediated reactions have been proposed (22, 59); each of these also features a single active site. The model described by Vink et al. (59) is similar to the one presented here in that there are two separate DNA-binding sites on each integrase monomer. In contrast, Engelman et al. (22) have suggested the possibility that there is only one DNA-binding site and the integration reaction is carried out by a tetramer in which one dimer binds the pair of viral DNA ends and the other dimer binds the target DNA. Many other possible models exist, but we favor the one depicted in Fig. 9 because it is the simplest one consistent with our results.

The reactions that depend absolutely on effective binding to viral DNA ends (3'-end processing, integration) were the most severely impaired by mutations in the HHCC region. Our previous work had shown that recognition of the disintegration substrate involves interactions with both the viral and target DNA components of the substrates (13, 14). However, the disintegration reaction promoted by wild-type integrase is much less sensitive to disruptions of the normal viral DNA end structure than are the 3'-end processing and integration reactions. For example, analyses of substrate requirements have shown that substitution of the CA dinucleotide severely impairs both processing and integration by HIV integrase (11, 35, 36, 49, 58). However, disintegration activity is reduced only sixfold by substitution of the CA dinucleotide with TC (14). These results may be interpreted to suggest that interactions with the target DNA portion of the Y substrate contribute substantially to substrate binding and thereby partially relieve the enzyme's dependence on effective interactions with the viral DNA component. By an analogous argument, the relative insensitivity of the disintegration reaction to disruption of the HHCC region may support a role for this region in recognition of the viral DNA substrate.

If the HHCC domain contributes to binding of the DNA substrate, then disrupting the remaining HHCC-independent protein-substrate interactions by altering the substrate structure might be expected to be more debilitating to the mutants than to wild-type integrase. Indeed, the HHCC mutants were all hypersensitive to modification of the target DNA component of the Y substrate (Fig. 6) and to unnatural replacement or structural alteration of the phylogenetically invariant CA-GT base pairs at the viral DNA end (Fig. 7 and 8). These results reinforce the conclusion that an intact HHCC region is not required for recognition of the target DNA or the invariant dinucleotide at the viral DNA end. Conversely, the heightened sensitivity of the mutant enzymes to these alterations implies that the HHCC region might play an important role in the residual interactions unaffected by these substrate modifications. Sequences internal to the invariant CA dinucleotide at the viral DNA end have previously been shown to affect recognition by integrase (35, 36, 48, 58). This subterminal portion of the viral DNA end is thus a leading candidate for a feature recognized by the HHCC region, since the enzyme's sensitivity to alterations in this part of the substrate was not enhanced by

mutation in the HHCC domain (Fig. 6A versus C and 8B versus C). In addition, the mutant and wild-type proteins are affected to similar extents by block substitutions of 5 to 6 bp in the viral DNA strand of the disintegration substrate (data not shown). This result, taken with those depicted in Fig. 6 to 8, suggests that recognition of the viral strand by the HHCC region is more dependent on the structure than on the sequence of the viral DNA. This recognition event may play a role in positioning the viral DNA end into the active site.

Although integration activity was not detected with mutants IN(H12A, E13L), IN(V32G, A33L, K34A), and IN(S39T, C40A) in the integration assay, there is some evidence that these mutants retain the ability to integrate under some conditions. A ladder of bands visible between the substrate and product is commonly observed in disintegration assays (Fig. 5A, 7A, and 8A). These are believed to arise from reintegration of the viral DNA strand following disintegration. The evidence for this is as follows. Similar minor products were observed when the Y substrate was labelled on the 33-mer; in addition to bands between the substrate and product, other bands that migrate more slowly than the input substrate appeared (Fig. 5B). By virtue of their size (unlike in the reactions in Fig. 5A), they cannot merely be nuclease degradation products; they probably arose from reintegration into either the viral strand of the disintegration substrate or the target strand, 5' of the original position. The prevalence of these larger species correlates with that of the bands visible between the substrate and product and with the level of integration activity observed for each mutant in the forward assay. Direct sequencing data also support the assertion that these are indeed reintegration products (data not shown). In addition, bands between the substrate and product did not appear when disintegration assays were performed with a substrate that has a CA-to-TC substitution, as might be expected if these are indeed due to reintegration. The fact that some HHCC mutants are able to accomplish reintegration suggests that they are deficient in a step leading to formation of an active primary complex with the viral DNA substrate and that this can be a limiting step in the reaction. Once formed, however, the complex has considerable stability and is fully active. By virtue of their remaining target DNA-binding activity, the mutants might be able to bind the disintegration substrate, form an active complex and, remaining bound to the viral portion, promote a subsequent integration event. The idea that formation of the initial complex with the viral DNA substrate is indeed a limiting step is further supported by the finding that mutant IN(P29A) (and also the wild-type protein) shows the same level of integration activity with the standard and preprocessed integration substrates (56).

While our results are most simply explained by a direct role of the HHCC region in substrate recognition, many alternative explanations are compatible with our data. An obvious alternative role for the HHCC domain might be in protein-protein interactions. Other zinc finger domains have been shown to play such a role (reviewed in references 4 and 30). As discussed above, protein-protein interactions may be important for both the 3'-end processing and integration activities of integrase. Thus, mutations in the HHCC region might be expected to interfere with both activities. Such a hypothetical role for the HHCC domain does not readily explain the altered sensitivity of the mutants to modification of the disintegration substrate, but those results could be accommodated by supposing, for example, that both protomers of an integrase dimer might contribute portions of each viral DNA-binding site.

Analyses of the DNA-binding activities of Rous sarcoma virus (33, 39) and HIV (46, 60) integrase mutants have generated results consistent with those described here. Deletion or substitution of the conserved histidines does not interfere with nonspecific DNA binding as measured either by DNA cellulose affinity chromatography (39), filter binding (33, 46), UV cross-linking (46), or Southwestern (60) assays. Binding in those assays was compromised by deletion of residues in the C-terminal half of integrase. Those results suggest that a target or nonspecific DNA-binding domain is likely to be located in the C-terminal region of the protein. Continued analysis of integrase by mutagenesis should identify residues directly involved in this and other critical aspects of integrase function, contributing to a thorough understanding of the mechanism of retroviral integration.

ACKNOWLEDGMENTS

We acknowledge the technical assistance of Stanley Brobston. We thank Andrew Leavitt and Harold Varmus for helpful discussions and Ronald Plasterk and Robert Craigie for communication of results prior to publication.

This work was supported by NIH grant AI27205 and by the Howard Hughes Medical Institute. P.O.B. is an assistant investigator of the Howard Hughes Medical Institute.

REFERENCES

- Adachi, A., H. E. Gendelman, S. Koenig, T. Folks, R. Willey, A. Rabson, and M. A. Martin. 1986. Production of acquired immunodeficiency syndrome-associated retrovirus in human and non-human cells transfected with an infectious molecular clone. *J. Virol.* **59**:284-291.
- Berg, J. M. 1986. Potential metal-binding domains in nucleic acid binding proteins. *Science* **232**:485-487.
- Berg, J. M. 1989. DNA binding specificity of steroid receptors. *Cell* **75**:1065-1068.
- Berg, J. M. 1990. Zinc fingers and other metal-binding domains. *J. Biol. Chem.* **265**:6513-6516.
- Bowerman, B., P. O. Brown, J. M. Bishop, and H. E. Varmus. 1989. A nucleoprotein complex mediates the integration of retroviral DNA. *Genes Dev.* **3**:469-478.
- Brown, P. O. 1990. Integration of retroviral DNA. *Curr. Top. Microbiol. Immunol.* **157**:19-48.
- Brown, P. O., B. Bowerman, H. E. Varmus, and J. M. Bishop. 1987. Correct integration of retroviral DNA *in vitro*. *Cell* **49**:347-356.
- Brown, P. O., B. Bowerman, H. E. Varmus, and J. M. Bishop. 1989. Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein. *Proc. Natl. Acad. Sci. USA* **86**:2525-2529.
- Brown, R. S., C. Sander, and P. Argos. 1985. The primary structure of transcription factor TFIIIA has 12 consecutive repeats. *FEBS Lett.* **186**:271-274.
- 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.
- Bushman, F. D., and R. Craigie. 1991. Activities of human immunodeficiency virus (HIV) integration protein *in vitro*: specific cleavage and integration of HIV DNA. *Proc. Natl. Acad. Sci. USA* **88**:1339-1343.
- Bushman, F. D., T. Fujiwara, and R. Craigie. 1990. Retroviral DNA integration directed by HIV integration protein *in vitro*. *Science* **249**:1555-1558.
- Chow, S. A., and P. O. Brown. Unpublished data.
- 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.
- Clavel, F., M. D. Hoggan, R. L. Willey, K. Stregel, M. A. Martin, and R. Repaske. 1989. Genetic recombination of human immunodeficiency virus. *J. Virol.* **63**:1455-1459.
- Colicelli, J., and S. P. Goff. 1985. Mutant and pseudo revertants of Moloney murine leukemia virus with alterations at the integration site. *Cell* **42**:573-580.
- Colicelli, J., and S. P. Goff. 1988. Sequence and spacing requirements of a retrovirus integration site. *J. Mol. Biol.* **199**:47-59.
- 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.
- Donehower, L. A. 1988. Analysis of mutant Moloney murine leukemia viruses containing linker insertion mutations in the 3' region of *pol*. *J. Virol.* **62**:3958-3964.
- Donehower, L. A., and H. E. Varmus. 1984. A mutant murine leukemia virus with a single missense codon in *pol* is defective in a function affecting integration. *Proc. Natl. Acad. Sci. USA* **81**:6461-6465.
- Ellison, V., H. Abrams, T. Roe, J. Lifson, and P. O. Brown. 1990. Human immunodeficiency virus integration in a cell-free system. *J. Virol.* **64**:2711-2715.
- 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.
- Evans, R. M., and S. M. Hollenberg. 1988. Zinc fingers: guilt by association. *Cell* **52**:1-3.
- Farnet, C., and W. A. Haseltine. 1990. Integration of human immunodeficiency virus type 1 DNA *in vitro*. *Proc. Natl. Acad. Sci. USA* **87**:4164-4168.
- Farnet, C., and W. A. Haseltine. 1991. Determination of viral proteins present in the human immunodeficiency virus type 1 preintegration complex. *J. Virol.* **65**:1910-1915.
- Fujiwara, T., and K. Mizuuchi. 1988. Retroviral DNA integration: structure of an integration intermediate. *Cell* **54**:497-504.
- Grandgenett, D. P., and S. R. Mumm. 1990. Unraveling retroviral integration. *Cell* **60**:3-4.
- Hemsley, A., N. Arnheim, M. D. Toney, G. Cortopassi, and D. J. Galas. 1989. A simple method for site-directed mutagenesis using the polymerase chain reaction. *Nucleic Acids Res.* **17**:6545-6551.
- 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 nonviral enzymes. *Proc. Natl. Acad. Sci. USA* **83**:7648-7652.
- 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.
- Kaptein, R. 1991. Zinc fingers. *Curr. Opin. Struct. Biol.* **1**:63-70.
- 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.
- 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.
- Khan, E., J. P. G. Mack, R. A. Katz, J. Kulkosky, and A. M. Skalka. 1991. Retroviral integrase domains: DNA binding and the recognition of LTR sequences. *Nucleic Acids Res.* **19**:851-860.
- Klug, A., and D. Rhodes. 1987. 'Zinc-fingers': a novel protein motif for nucleic acid recognition. *Trends Biochem. Sci.* **12**:464-469.
- LaFemina, R. L., P. L. Callahan, and M. G. Cordingley. 1991. Substrate specificity of recombinant human immunodeficiency virus integrase protein. *J. Virol.* **65**:5624-5630.
- 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.

37. Leavitt, A. D., L. Shiue, and H. E. Varmus. (University of California, San Francisco). 1992. Personal communication.
38. Miller, J., A. D. McLachlan, and A. Klug. 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J.* **4**:1609–1614.
39. Mumm, S. R., and D. P. Grandgenett. 1991. Defining nucleic acid-binding properties of avian retrovirus integrase by deletion analysis. *J. Virol.* **65**:1160–1167.
40. Murphy, J., and S. P. Goff. 1992. A mutation at one end of Moloney murine leukemia virus DNA blocks cleavage of both ends by the viral integrase in vivo. *J. Virol.* **66**:5092–5095.
41. Panganiban, A. T., and H. M. Temin. 1984. The retrovirus *pol* gene encodes a product required for DNA integration: identification of a retrovirus *int* locus. *Proc. Natl. Acad. Sci. USA* **81**:7885–7889.
42. Parraga, G., S. Horvath, L. Hood, E. T. Young, and R. E. Klevit. 1990. Spectroscopic studies of wild-type and mutant “zinc-finger” peptides: determinants of domain folding and structure. *Proc. Natl. Acad. Sci. USA* **87**:137–141.
43. Roth, M. J., P. L. Schwartzberg, and S. P. Goff. 1989. Structure of the termini of DNA intermediates in the integration of retroviral DNA. Dependence on IN function and terminal DNA sequence. *Cell* **58**:47–54.
44. Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**:487–491.
45. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
46. Schauer, M., and A. Billich. 1992. The N-terminal region of HIV-1 integrase is required for integration activity, but not for DNA binding. *Biochim. Biophys. Acta* **186**:874–888.
47. Schwartzberg, P., J. Colicelli, M. L. Gordon, and S. P. Goff. 1984. Mutations in the *gag* gene of Moloney murine leukemia virus: effects on production of virions and reverse transcriptase. *J. Virol.* **49**:918–924.
48. Sherman, P. A., M. L. Dickson, and J. A. Fyfe. 1992. Human immunodeficiency virus type 1 integration protein: DNA sequence requirements for cleaving and joining reactions. *J. Virol.* **66**:3593–3601.
49. 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.
50. Siegel, L. M., and K. J. Monty. 1966. Determination of molecular weights and frictional ratios of proteins in impure systems by use of gel filtration and density gradient centrifugation. Application to crude preparations of sulfite and hydroxylamine reductases. *Biochim. Biophys. Acta* **112**:346–362.
51. Stevenson, M., S. Haggerty, C. A. Lamonica, C. M. Meier, S.-K. Welch, and A. J. Wasiaak. 1990. Integration is not necessary for expression of human immunodeficiency virus type 1 protein products. *J. Virol.* **64**:2421–2425.
52. 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.
53. van Gent, D. C., Y. Elgersma, M. W. J. Bolk, C. Vink, and R. H. A. Plasterk. 1991. DNA binding properties of the integrase proteins of human immunodeficiency viruses types 1 and 2. *Nucleic Acids Res.* **19**:3821–3827.
54. 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.
55. 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.
56. Vincent, K. A., and P. O. Brown. Unpublished data.
57. Vincent, K. A., D. York-Higgins, M. Quiroga, and P. O. Brown. 1991. Host sequences flanking the HIV provirus. *Nucleic Acids Res.* **18**:6045–6047.
58. Vink, C., D. C. van Gent, Y. Elgersma, and R. H. A. Plasterk. 1991. Human immunodeficiency virus integrase protein requires a subterminal position of its viral DNA recognition sequence for efficient cleavage. *J. Virol.* **65**:4636–4644.
59. Vink, C., E. Yeheskiely, G. A. van der Marel, J. H. van Boom, and R. H. A. Plasterk. 1991. Site-specific hydrolysis and alcoholysis of human immunodeficiency virus DNA termini mediated by the viral integrase protein. *Nucleic Acids Res.* **19**:6691–6698.
60. Woerner, A. M., M. Klutch, J. G. Levin, and C. J. Marcus-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.