JOURNAL OF VIROLOGY, Nov. 1992, p. 6361–6369
0022-538X/92/116361-09$02.00/0

Vol. 66, No. 11

Identification of Conserved Amino Acid Residues Critical for Human Immunodeficiency Virus Type 1 Integrase Function In Vitro

ALAN ENGELMAN AND ROBERT CRAIGIE*

Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, Maryland 20892

Received 17 June 1992/Accepted 3 August 1992

We have probed the structural organization of the human immunodeficiency virus type 1 integrase protein by limited proteolysis and the functional organization by site-directed mutagenesis of selected amino acid residues. A central region of the protein was relatively resistant to proteolysis. Proteins with altered amino acids in this region, or in the N-terminal part of the protein that includes a putative zinc-binding motif, were purified and assayed for 3′ processing, DNA strand transfer, and disintegration activities in vitro. In general, these mutations had parallel effects on 3′ processing and DNA strand transfer, suggesting that integrase may utilize a single active site for both reactions. The only proteins that were completely inactive in all three assays contained mutations at conserved amino acids in the central region, suggesting that this part of the protein may be involved in catalysis. In contrast, none of the mutations in the N-terminal region resulted in a protein that was inactive in all three assays, suggesting that this part of integrase may not be essential for catalysis. The disintegration reaction was particularly insensitive to these amino acid substitutions, indicating that some function that is important for 3′ processing and DNA strand transfer may be dispensable for disintegration.

Integration of a DNA copy of the retroviral RNA genome into a chromosome of the host cell is necessary for efficient virus replication. Retroviral DNA integration is accomplished by a defined set of DNA cutting and joining reactions mediated by the viral integration machinery. In the first step, two nucleotides are removed from the 3′ ends of the linear viral DNA made by reverse transcription. The resulting recessed 3′ ends of the viral DNA are then inserted into the target DNA in a subsequent DNA strand transfer reaction; the 5′ ends of the viral DNA and the 3′ ends of target DNA at the site of insertion remain unjoined in the resulting integration intermediate. The integration process is completed by degradation of the two unpaired bases at the 5′ ends of the viral DNA and repair of the single-strand gaps between viral and target DNA, reactions that are likely to be carried out by host enzymes. See references 3 and 18 for reviews of retroviral DNA integration.

Purified human immunodeficiency virus type 1 (HIV-1) integrase (IN) protein possesses all biochemical activities necessary for formation of the integration intermediate. Integrase has a site-specific nuclease activity that cleaves two nucleotides from the 3′ end of oligonucleotide DNA substrates that model the ends of HIV DNA (5, 32, 48, 56). This 3′ processing reaction exposes the CA_{OH}-3′ end that is inserted into a target DNA in a subsequent DNA strand transfer reaction that is also carried out by integrase (5, 6, 32). The integrase proteins of several other retroviruses, including Moloney murine leukemia virus (Mo-MLV) (9) and avian sarcoma-leukosis virus (27, 28, 58), also mediate these reactions in vitro. HIV-1 integrase has also been shown to promote an apparent reversal of the DNA strand transfer reaction, termed disintegration (7).

These in vitro systems using oligonucleotide substrates and purified integrase protein have been invaluable for investigating the DNA substrate requirements (5, 32, 33, 55, 56) and chemical mechanisms (13, 57) of the 3′ processing and DNA strand transfer reactions catalyzed by HIV-1 integrase. The stereochemical course of both the 3′ processing and DNA strand transfer reactions has been monitored by incorporating phosphorothioate of known chirality in the substrate DNA for in vitro reactions and determining the chirality of these groups in the reaction products (13). Phosphorothioate at the site of cleavage undergoes inversion of chirality during the 3′ processing reaction. Similarly, the chirality of phosphorothioate at the site of insertion in the target DNA is inverted during the course of the DNA strand transfer reaction. These results strongly support the view that these reactions both proceed by a one-step mechanism not involving a covalent intermediate between integrase and DNA (13).

Although progress has been made in determining the DNA substrate requirements for integrase and understanding the chemical mechanism of the 3′ processing and DNA strand transfer reactions, little is known about the parts of the protein that are involved in interactions with the DNA substrates, protein-protein interactions, or catalysis. We have used limited proteolysis to probe the structure of HIV-1 integrase and site-directed mutagenesis of selected conserved amino acid residues (14, 26, 29, 31) to address the functional organization of the protein. The results suggest that integrase may have a single active site located in a central domain of the protein that is relatively resistant to proteolysis.

MATERIALS AND METHODS

Nucleotides, oligonucleotides, and enzymes. Radionucleotides were obtained from New England Nuclear at a specific activity of 3,000 Ci/mmol. Oligonucleotides were synthesized by using an Applied Biosystems 380B DNA synthesizer. Restriction endonucleases and T4 DNA ligase were obtained from New England Biolabs. T4 polynucleotide kinase was obtained from Pharmacia LKB. Sequenase

* Corresponding author.

version 2.0 DNA polymerase was obtained from United States Biochemical Corp. Proteases were obtained from Boehringer Mannheim.

Construction of integrase expression plasmids. HIV-1 integrase was expressed in *Escherichia coli*, essentially as described previously (48), using a T7 expression system (53). Plasmid pMK561 (6) was digested with *Bsp*MI, and the 741-bp DNA fragment that includes most of the integrase coding region was isolated. Double-stranded oligonucleotides were ligated onto each end of the DNA fragment to restore the complete coding region. One oligonucleotide restored the 5′ end of the region and included both an ATG codon for initiation of translation in *E. coli* and an *Nde*I cohesive end. Another oligonucleotide restored the 3′ end of the region and included a *Bam*HI cohesive end. This DNA was ligated with *Nde*I-*Bam*HI-digested pAR2156 DNA (42). The resulting plasmid (pINT7) was digested with *Xba*I and *Bam*HI, and the 911-bp DNA fragment that includes the entire integrase coding sequence and some 5′-proximal vector sequence was isolated and ligated with *Xba*I-*Bam*HI-digested M13mp19 DNA. Single-stranded phage DNA was prepared, and bases 4673 and 4676 of the integrase coding region were altered by site-directed mutagenesis to introduce previously determined substitutions that decrease the initiation of translation from an adventitious internal start site (24). Mutagenesis was performed by using a kit obtained from Amersham (catalog no. RPN 1523). The entire integrase coding region in the resulting recombinant phage (B2) was sequenced by the dideoxynucleotide chain termination method (45). Phage B2 was used to prepare double-stranded DNA, and this DNA was digested with *Xba*I and *Bam*HI. The 911-bp DNA fragment was isolated, and plasmid pINSD was produced by ligation of this fragment with *Xba*I-*Bam*HI-digested pAR2156 DNA.

Specific codons of the integrase coding region were mutagenized by using oligonucleotides and single-stranded DNA from phage B2. The presence of the desired base changes was confirmed by sequencing the targeted region of products of each mutagenesis reaction. The 911-bp *Xba*I-*Bam*HI DNA fragments containing the desired mutations were ligated with *Xba*I-*Bam*HI-digested pAR2156 DNA to obtain expression plasmids for mutant integrase proteins.

Protein purification. HIV-1 integrase was purified from *E. coli* essentially as described previously (48), with some minor modifications. *E. coli* BL21(DE3) (53) was transformed with an integrase expression plasmid, plated onto agar plates containing 100 μg of ampicillin per ml, and incubated at 37°C overnight. A colony was inoculated into Super broth (Quality Biological) containing 100 μg of ampicillin per ml, and the culture was grown to an optical density at 600 nm of 0.8. Glycerol was added to a final concentration of 15% (wt/vol), and the culture was frozen at −70°C. For induction of integrase expression, the cells were thawed and diluted 1:100 (vol/vol) into 5 liters of Super broth containing 100 μg of ampicillin per ml. The culture was grown to an optical density at 600 nm of 0.8, and isopropylthiogalactopyranoside (IPTG) was added to a final concentration of 0.4 mM. Incubation was continued for 3 h, and the cells were harvested and resuspended in 120 ml of ice-cold 25 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid (HEPES; pH 7.6)–1 mM EDTA. The cells were frozen in liquid nitrogen. The rest of the preparation was done at 4°C or on ice.

The cells were thawed on ice overnight and resuspended in an equal volume of LS buffer (0.1 M NaCl, 20 mM Tris-HCl [pH 8.0], 1 mM EDTA, 1 mM dithiothreitol [DTT]), and lysozyme was added to a final concentration of 0.2 mg/ml. The cells were incubated for 30 min on ice and then subjected to six cycles of sonication for 20 s at 100 W, allowing 4 min for cooling between each cycle. The lysed cells were centrifuged at 30,000 × *g* for 25 min, and the pellet was resuspended in 40 ml of LS buffer by homogenization. The suspension was centrifuged at 30,000 × *g* for 25 min, and the pellet, which contained the integrase, was resuspended in 40 ml of buffer containing 1 M NaCl, 20 mM Tris-HCl (pH 7.5), 1 mM EDTA, and 1 mM DTT by homogenization. The suspension was stirred slowly for 30 min and then centrifuged at 30,000 × *g* for 25 min. Ammonium sulfate (0.8 M, final concentration) was added to the supernatant, which contained the solubilized integrase.

The extract was loaded onto 2 ml of butyl-Sepharose 4B (Pharmacia LKB). The column was equilibrated prior to loading by being washed first with LSC buffer (50 mM HEPES [pH 7.6], 0.2 M NaCl, 0.1 mM EDTA, 1 mM DTT, 10% [wt/vol] glycerol) and then with HSC buffer (50 mM HEPES [pH 7.6], 0.2 M NaCl, 0.8 M ammonium sulfate, 0.1 mM EDTA, 1 mM DTT). Proteins were eluted by using a linear gradient of 100% HSC buffer to 100% LSC buffer. The fractions containing integrase were pooled and diluted 1:2 (vol/vol) with 50 mM HEPES (pH 7.6)–0.1 mM EDTA–1 mM DTT–10% (wt/vol) glycerol. This material was loaded onto 1 ml of heparin-Sepharose CL-4B (Pharmacia LKB). The column was washed first with HS buffer (50 mM HEPES [pH 7.6], 1 M NaCl, 0.1 mM EDTA, 1 mM DTT, 10% [wt/vol] glycerol) and then with LSC buffer prior to loading. Proteins were eluted with a linear gradient of 100% LSC buffer to 100% HS buffer. Integrase-containing fractions were pooled and dialyzed against 20 mM HEPES (pH 7.6)–1 M NaCl–1 mM EDTA–1 mM DTT–20% (wt/vol) glycerol. The proteins were further purified by dialysis against 20 mM HEPES (pH 7.6)–0.4 M potassium glutamate–0.1 mM EDTA–1 mM DTT–0.1% (vol/vol) Nonidet P-40, 20% (wt/vol) glycerol, which precipitates integrase, leaving most of a contaminating nuclease activity in the supernatant (12). The resulting suspension was centrifuged at 12,500 × *g* for 25 min, and the pellet was resuspended in 20 mM HEPES (pH 7.6)–1 M NaCl–1 mM EDTA–1 mM DTT–10% (wt/vol) glycerol. This mixture was incubated for 30 min and then centrifuged at 12,500 × *g* for 25 min. The supernatant, which contained the soluble integrase, was frozen in liquid nitrogen and stored at −70°C.

Proteolysis and protein sequencing. HIV-1 integrase was subjected to proteolysis with three different proteases: trypsin, chymotrypsin, and V8 protease. Approximately 2 μg of integrase was incubated with 40 ng of protease in 22 μl of 20 mM HEPES (pH 7.6)–0.55 M NaCl–1 mM EDTA–1 mM DTT–20% (wt/vol) glycerol. Reactions with V8 protease were carried out at 25°C, and reactions with trypsin and chymotrypsin were carried out at 30°C. Reactions were stopped by the addition of phenylmethylsulfonyl fluoride to a final concentration of 2 mM followed by the addition of 5.5 μl of 5× protein sample buffer (0.31 M Tris-HCl [pH 6.8], 10% [wt/vol] sodium dodecyl sulfate [SDS], 50% [wt/vol] glycerol, 25% [vol/vol] 2-mercaptoethanol, 0.005% [wt/vol] bromophenol blue).

Reactions were scaled up fivefold for protein sequencing. The products of proteolysis were electrophoresed in SDS–15% polyacrylamide (30:0.8, acrylamide-bisacrylamide) gels and transferred to polyvinylidene difluoride membranes as described previously (36). The membranes were stained with Coomassie blue R-250, and stained bands were cut from the membrane and subjected to 20 cycles of Edman degradation, using an Applied Biosystems 471A protein sequencer.

Protein sequence alignment. The predicted amino acid

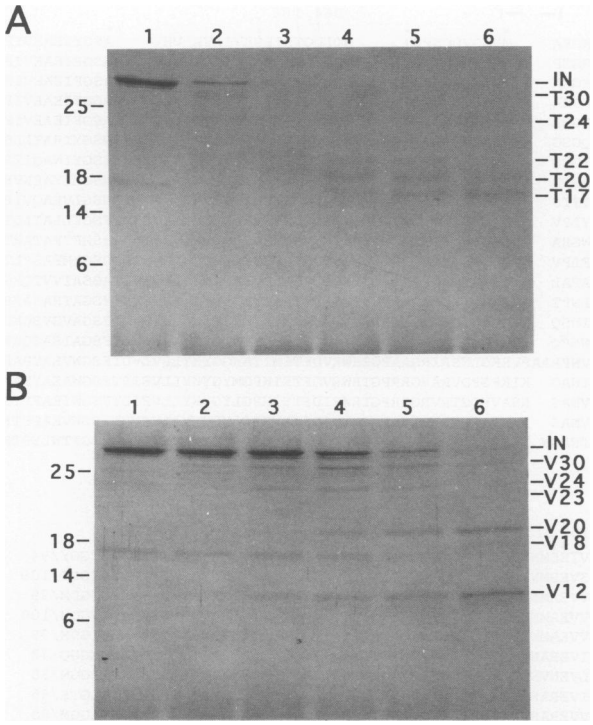

FIG. 1. Limited proteolysis of HIV-1 integrase with trypsin and V8 protease. (A) Integrase was incubated with trypsin, and the cleavage products were analyzed on an SDS–15% polyacrylamide gel. Lanes: 1, phenylmethylsulfonyl fluoride was added prior to trypsin; 2 to 6, reactions for 0.5, 2, 4, 8, and 20 h, respectively. The migration positions of intact integrase and the major proteolytic fragments are indicated on the right; the migration positions of molecular weight standards are indicated in kilodaltons on the left. (B) Limited proteolysis with V8 protease. The reactions in lanes 1 to 6 were the same as in panel A except that V8 protease was substituted for trypsin; the major proteolytic fragments are labeled.

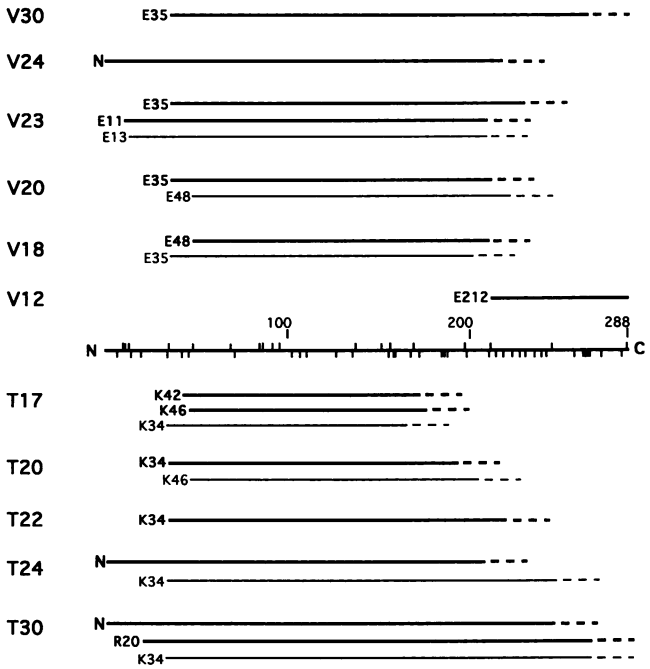

FIG. 2. Locations of the protease cleavage sites. Intact integrase is represented as a bold line (center). Amino acid positions are numbered from the N terminus. The positions of Glu residues (V8 cleavage sites) and Arg and Lys residues (trypsin cleavage sites) are indicated above and below, respectively. The peptides generated by V8 cleavage are shown above, and those generated by cleavage with trypsin are shown below. Predominant peptides are noted in bold type when the band on the gel contained a mixture of peptides that were not resolved by electrophoresis. Dashed lines represent the estimated positions of C termini.

sequences of the integrase proteins from 20 retroviruses were aligned with the amino-terminal 194 amino acids of HIV-1 integrase from strain NL43 (1, 34). Sequences were aligned by eye to yield maximum similarity.

3′ processing, strand transfer, and disintegration reactions. The following oligonucleotides were used as DNA substrates: AE117 (5′-ACTGCTAGAGATTTTCCACAC), AE119 (5′-GTGTGGAAAATCTCTAGCA), AE146 (5′-GG ACGCCATAGCCCCGGCGCGGTCGCTTTC), AE150 (5′-GTGTGGAAAATCTCTAGCAG), AE156 (5′-GTGTGGAA AATCTCTAGCAGGGGCTATGGCGTCC), and AE157 (5′-GAAAGCGACCGCGCC).

Oligonucleotides AE117 and AE150 were annealed and filled in with [α-^{32}P]TTP, using Sequenase version 2.0 DNA polymerase for use in the 3′ processing reaction as described previously (13). Processing reactions and denaturing polyacrylamide gel electrophoresis were also done as described previously (13).

The preprocessed substrate for the DNA strand transfer reaction was prepared by labeling oligonucleotide AE119 with T4 polynucleotide kinase and annealing an equimolar amount of oligonucleotide AE117 as previously described (9). Labeled substrate was separated from unincorporated nucleotides by passage through a G25 quick-spin column (Boehringer Mannheim). Conditions for the DNA strand transfer reaction were identical to those previously described for the 3′ processing reaction (13).

The DNA substrate for the disintegration reaction was prepared by labeling oligonucleotide AE157 with T4 polynucleotide kinase and annealing equimolar amounts of oligonucleotides AE146, AE117, and AE156 in the same way as described for the preparation of the DNA strand transfer substrate. This substrate was also passed through a G25 spin column prior to use. Conditions for the disintegration reaction were identical to those for the 3′ processing reaction.

Quantitation of protein concentration and reaction products. Protein concentrations were determined by densitometry of bands in SDS-polyacrylamide gels stained with Coomassie blue G-250, with wild-type integrase as a standard, using a MasterScan Interpretive Densitometer (Scanalytics).

Products of the 3′ processing, DNA strand transfer, and disintegration reactions were quantitated by using a Molecular Dynamics PhosphorImager.

RESULTS

Mapping of accessible protease cleavage sites in HIV-1 integrase. Partial proteolysis is a useful tool for analyzing the domain structure of proteins; protease cleavage sites located within tightly folded regions of a protein are more resistant to cleavage than are sites located in hinge regions between domains or in other, less ordered parts of the structure. We have subjected HIV-1 integrase to limited proteolysis and analyzed the cleavage products by SDS-polyacrylamide gel electrophoresis and N-terminal peptide sequencing.

A time course of digestion with trypsin and V8 protease is

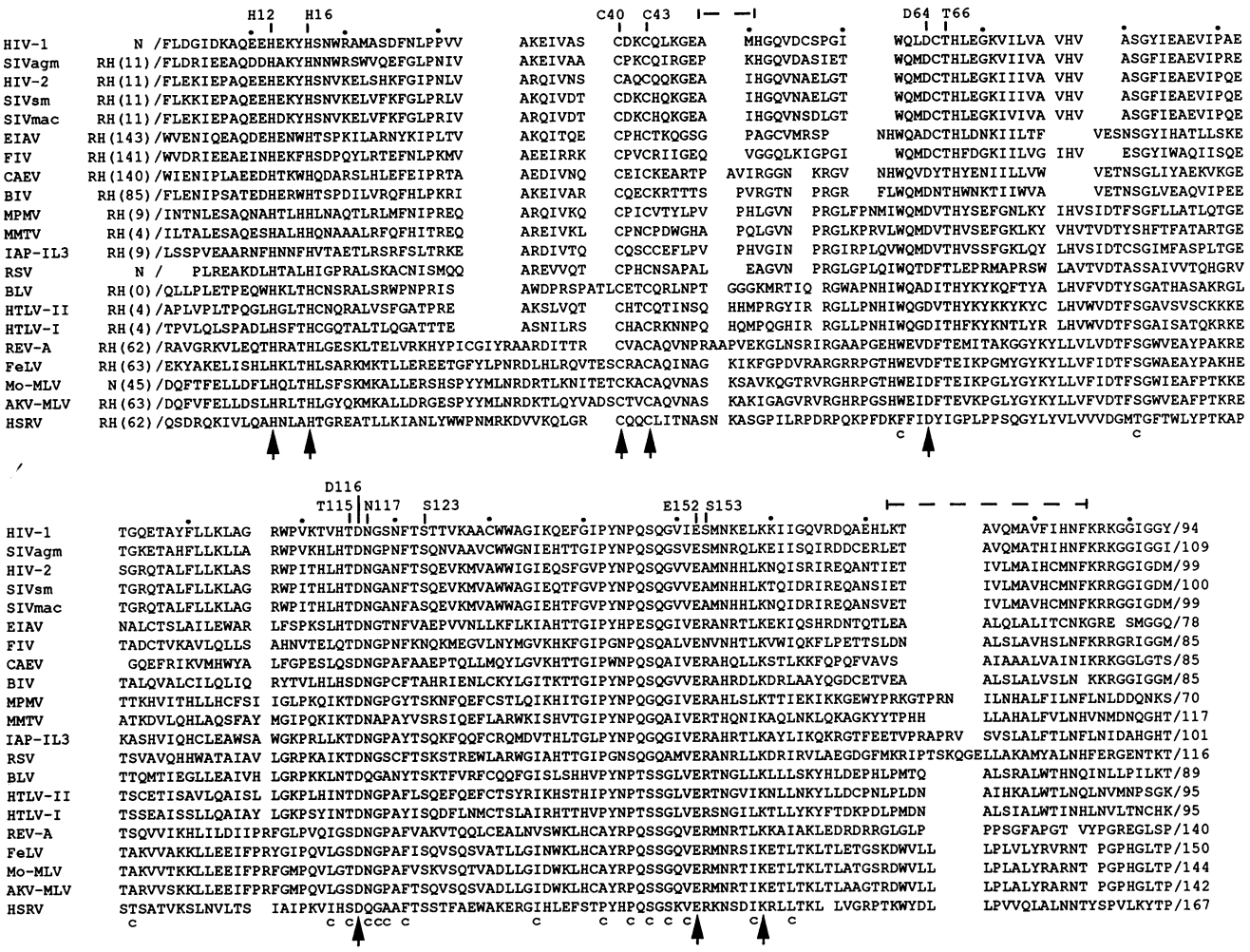

FIG. 3. Protein sequence alignment. The amino acid sequence of HIV-1 integrase, from strain NL43, is shown from the amino terminus to residue 194 (1, 34). The dots above the sequence number every 10 residues and the dashed lines mark the approximate boundary of the central core. The residues highlighted above the sequence were changed by site-directed mutagenesis. The symbols below the sequences indicate positions of amino acid conservation (arrow, invariant amino acid residue; c, amino acid residue that exhibits only conservative variation). Amino acid residues were grouped for this purpose as follows: G, A, S, T, P; L, I, V, M; F, Y, W; D, E, N, Q; K, R, H; C (47). Numbers at the left are the numbers of amino acid residues between the amino termini and the conserved Asp residue in the RNase H (RH) domains of the viruses. This Asp is residue 704 of the p61 reverse transcriptase protein of HIV-1 NL43 (1). For Mo-MLV, the number at the left is the number of residues between the terminus and the presumed amino terminus of Mo-MLV integrase protein (8). N/, HIV-1 and Rous sarcoma virus integrase N termini (22, 34). Numbers at the right are the numbers of amino acid residues remaining in the complete integrase proteins. The other sequences were aligned to the HIV-1 sequence by eye to yield maximum similarity. These sequences are from the following viruses: SIVagm, simian immunodeficiency virus from African green monkey, strain 677 (16); HIV-2, HIV type 2, strain ROD (19); SIVsm, SIV from a sooty mangabey, strain H4 (23); SIVmac, SIV from a macaque, strain 251 (30); EIAV, equine infectious anemia virus, strain CL22 (52, 59); FIV, feline immunodeficiency virus, strain Petaluma (54); CAEV, caprine arthritis-encephalitis virus, strain CO (44); BIV, bovine immunodeficiency virus, clone 127 (17); MPMV, Mason-Pfizer monkey virus, strain 6A (51); MMTV, mouse mammary tumor virus, BR6 isolate (38); IAP-IL3, intracisternal A particle, IL3 isolate (20, 61); RSV, Rous sarcoma virus, strain Prague C (46); BLV, bovine leukemia virus, strain FLK (41); HTLV-II, human T-cell leukemia virus type II, clone λH6.0 (49); HTLV-I, human T-cell leukemia virus type I, clone HS-35 (35); REV-A, reticuloendotheliosis virus, strain A (60); FeLV, feline leukemia virus, strain F6A (10); Mo-MLV, clone pMLV-1 (37, 50); AKV-MLV, murine leukemia virus, strain AKR-623 (21); HSRV, human spumaretrovirus, clone pHSRV (15, 40).

shown in Fig. 1; chymotrypsin gave a series of proteolytic fragments similar to those generated by trypsin (12). The amino termini of the five major fragments observed during the time course of digestion with trypsin (Fig. 1A) and of the six fragments produced by digestion with V8 protease (Fig. 1B) were determined by peptide sequencing. The approximate carboxyl termini of these fragments were then deduced from their apparent molecular weights. The results of this analysis are summarized in Fig. 2.

The amino-terminal region of the protein was relatively sensitive to proteolysis; cleavage was detected at most of the predicted (1, 4, 25) cleavage sites for trypsin and V8 protease (Fig. 2). The carboxyl-terminal part of the protein was also relatively sensitive to proteolysis. In contrast, a central core region of the protein, comprising approximately 120 amino acid residues, was more resistant; proteolytic fragments spanning this region remained intact at late times during proteolysis (Fig. 1, fragments T17, T20, and V20).

Selection of amino acid residues for site-directed mutagenesis. HIV-1 integrase contains certain amino acid residues that are invariant among the sequences of retroviral and retrotransposon integrases, as well as the transposases of some bacterial transposable elements (14, 26, 29, 31, 43). Figure 3 shows an alignment of the amino acid sequence of HIV-1 integrase with the corresponding sequences from 20 other retroviruses; most of the sequences in this alignment are from infectious molecular clones of viral DNA and are therefore expected to encode functional integrase. We used this alignment to determine where the invariant residues lie with respect to the regions of the protein that were both relatively sensitive and relatively resistant to proteolysis. The alignment also identifies less well conserved amino acid residues among retroviral integrases that might be attractive targets for mutagenesis.

Eight amino acid residues are invariant in this sequence alignment (Fig. 3). Four of these residues are in the central core, and the other four are in the amino-terminal region that was relatively sensitive to proteolysis. In addition, 17 amino acids, all of which lie in the central core, show only conservative variation in amino acid side chains (Fig. 3).

Site-directed mutagenesis was used to introduce amino acid substitutions at selected invariant residues and some less conserved residues. These residues are noted above the sequence in Fig. 3 (see also Table 1). Three of the four invariant residues in the central core, D-64, D-116, and E-152, are also conserved in the sequences of retrotransposons and certain bacterial transposable elements (14, 31, 43). These amino acids were replaced by residues with either a different carboxyl side chain or an amide side chain. We also made conservative substitutions in the four invariant residues in the amino-terminal region. These His and Cys residues are conserved in the corresponding regions of retrotransposons but not in the sequences of bacterial transposable elements (29, 43). Five other residues of the central core were also changed by mutagenesis. In general, these residues were chosen because they were both highly conserved in our alignment and near one of the targeted invariant residues (Fig. 3). In addition, Ala was substituted for S-123 as a control; Ala is present at this position in four of the other sequences in the alignment, and Gly is present at this position in an independent infectious molecualr clone of HIV-1 (39).

Biochemical activities of the mutant integrase proteins. Each of the mutant proteins was expressed in *E. coli* and purified. We chose to study only those mutant proteins that could be purified by the procedure used for the wild-type protein and are therefore unlikely to have gross structural changes induced by the point mutations. Some additional mutants, with altered purification properties, were not studied.

The 3′ processing, DNA strand transfer, and disintegration activities of each mutant protein were assayed over a range of protein concentration. The concentration dependence of each activity with the wild-type protein is shown in Fig. 4. Each of the mutant proteins was assayed in the same way for the three activities. Figures 5 and 6 shows the results for each mutant at an integrase concentration of 0.30 μM; any differences in concentration dependence between reactions with the mutant and wild-type proteins are noted in the text. The data are summarized in Table 1.

We analyzed the activities of four mutant proteins containing amino acid substitutions in the His-Cys region. Two of these proteins contained single amino acid substitutions of either H-12 or C-40; the other two contained substitutions of either both His or both Cys residues. The H12N substitution decreased both 3′ processing and DNA strand transfer about threefold in comparison with the wild-type protein, but disintegration activity was unaffected (Fig. 5 and 6; Table 1). The double substitution H12N/H16N had a much more drastic effect. The 3′ processing and DNA strand transfer activities of this protein were below the limits of detection, with reductions in activity of at least 300-fold (Fig. 5 and 6A; Table 1). However, a decrease of only about fivefold was observed in the disintegration activity (Fig. 6B). Both the 3′ processing and DNA strand transfer activities remained undetectable when wild-type and H12N/H16N proteins were compared in reactions containing a ninefold-higher concentration of protein and the same concentration of DNA substrate (12). The single C40S substitution affected disintegration activity to a similar extent as did substitution of both His residues (Fig. 6B and Table 1). However, unlike H12N/H16N, the C40S protein exhibited detectable 3′ processing and DNA strand transfer activities, albeit much reduced relative to that of the wild-type protein (Table 1). Substitution of both C-40 and C-43 did not further diminish any of the activities; in fact, this protein seemed to be somewhat more active and displayed disintegration activity similar to that of the wild-type protein (Fig. 6B; Table 1).

TABLE 1. Relative activities of HIV-1 IN proteins

Protein	Relative activity[a]		
	3′ processing	DNA strand transfer	Disintegration
H12N	30	38	135
H12N/H16N	<0.3	<0.1	18
C40S	1.2	3.8	19
C40S/C43S	1.9	5.5	64
D64N	<0.3	0.5	0.3
D64E	<0.3	0.6	0.3
T66A	22	42	91
T115A	80	95	140
D116N	<0.3	<0.1	<0.1
D116E	1.7	4.8	38
N117Q	26	57	99
S123A	71	68	136
E152D	<0.3	1.8	7.2
E152Q	<0.3	<0.1	<0.1
S153R	24	22	48

[a] Percent of wild-type activity at a protein concentration of 0.30 μM. Values are for the experiments shown in Fig. 5 and 6. In a separate set of assays, the same results were obtained within a tolerance of ±10% (12).

In contrast to the relative tolerance of the His-Cys region to mutation, two of the substitutions in the central core region, D116N and E152Q, completely abolished detectable activity in all three assays; substitution at these positions with a residue that conserves the carboxyl side chain resulted in proteins with partial activities (Table 1). Substitution of the other invariant residue in this region also drastically affected the activities of the protein; 3′ processing activity was undetectable with D64N and D64E, and DNA strand transfer and disintegration activities were only barely detectable with these mutants.

Substitution of the less well conserved residues in the central core region had lesser effects on activity (Fig. 5 and 6; Table 1). Reactions with the N117Q protein, which is quite active in all three assays, showed an interesting difference in the distribution of 3′ processing cleavage products (Fig. 5). Under our reaction conditions (13), these products include the simple dinucleotide, a cyclic form of this dinucleotide (labeled FORM II in Fig. 5), and an adduct in which glycerol

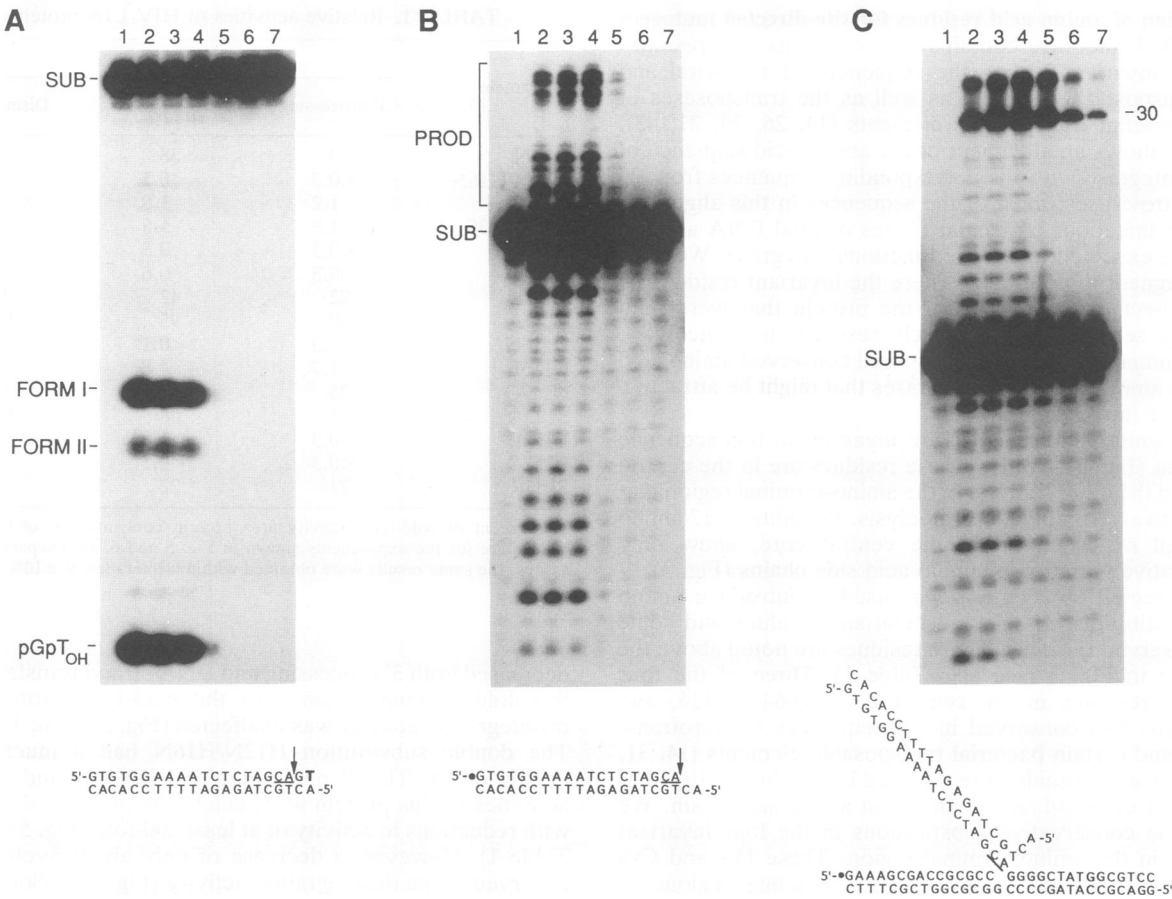

FIG. 4. 3′ processing, DNA strand transfer, and disintegration activities of wild-type HIV-1 integrase. The DNA substrates are shown below each panel. (A) 3′ processing reaction. The substrate models the terminal 21 bp of the U5 end of the linear viral DNA. The arrow depicts the site of cleavage by integrase. The DNA was labeled with ^{32}P at the phosphate 5′ of the T highlighted in bold type. Integrase was omitted from the reaction in lane 1; the reactions in lane 2 to 7 contained 2.3 μM, 0.77 μM, 0.26 μM, 85 nM, 26 nM, and 10 nM integrase, respectively. The migration position of the substrate DNA is marked SUB. Different nucleophiles can participate in the processing reaction mediated by integrase, giving rise to alternative products (13). FORM I is the product of cleavage by glycerol, FORM II is the product of cleavage by the 3′-OH of the DNA substrate, and pGpT$_{OH}$ is the product of hydrolysis (13, 57). (B) DNA strand transfer reaction. The arrow depicts the 3′ end of the preprocessed substrate that is joined to target DNA in the strand transfer product. The reactions in lanes 1 to 7 contained the same amounts of integrase as did the corresponding lanes in panel A. The migration position of the substrate DNA is marked SUB, and the products of strand transfer are marked PROD. (C) Disintegration reaction. The reactions in lanes 1 to 7 contained the same amounts of integrase as did the corresponding lanes in panel A. The migration position of the starting substrate DNA is marked SUB, and the disintegration product is marked 30. The smear above the reaction product is reportedly due to secondary structure of the product DNA (7). The conserved CA dinucleotide that marks the 3′ end of the viral sequence that is joined to target DNA in the strand transfer reaction is underlined in each panel. The labeled phosphates in panels B and C are marked (•). The 5′ end of each single-stranded oligonucleotide is noted in each panel.

is attached to the dinucleotide cleavage product (FORM I in Fig. 5). Whereas the glycerol adduct is more abundant than the cyclic product with the wild-type protein, the cyclic dinucleotide is the more abundant product with N117Q.

DISCUSSION

We have attempted to gain insight into the functional organization of HIV-1 integrase by analyzing the 3′ processing, DNA strand transfer, and disintegration activities of mutant proteins containing amino acid substitutions at conserved amino acid positions. The results support the idea that these activities are mediated by a common catalytic site, located within a central core region that includes three highly conserved amino acid residues, D-64, D-116, and E-152. The N-terminal part of the protein, which contains a putative zinc-binding motif, appears more tolerant to mutagenesis with respect to the three activities that we have assayed in vitro and may not be directly involved in catalysis.

The His-Cys region. Although the His-Cys motif is highly conserved throughout retroviruses, none of the mutations that we tested in this region reduced disintegration activity to an undetectable level, and only one of the four mutants, H12N/H16N, was scored as completely inactive in the 3′ processing and DNA strand transfer assays. Thus, although changes in these conserved residues can significantly influence the efficiency of the in vitro reactions, they do not abolish catalytic activity. The disintegration activity was particularly insensitive to changes in these residues; substitution of both His or both Cys residues decreased disintegration activity only about fivefold. Similarly, substitution of either of the corresponding His residues of Rous sarcoma virus integrase does not significantly impair 3′ processing or DNA strand transfer activity (29). The ability of proteins

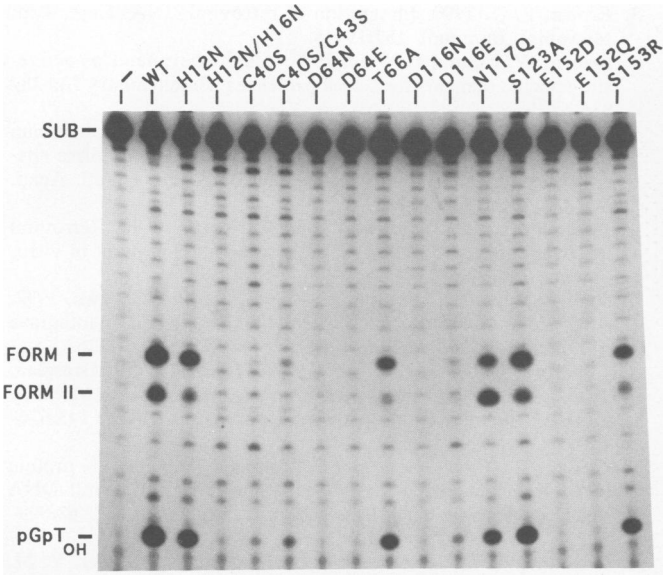

FIG. 5. 3′ processing reactions of mutant integrase proteins. Each protein was assayed for 3′ processing activity at a concentration of 0.30 μM. Reaction products were analyzed on a denaturing polyacrylamide gel. Integrase was omitted from the reaction in the lane marked −; the wild-type protein was added to the reaction in the lane marked WT; the other lanes contained the products of reactions with the indicated mutant proteins. Other labeling is the same as in Fig. 4A.

with substitutions at the conserved His and Cys residues to function in vitro suggests that the His-Cys motif is not essential for catalysis. It is tempting to speculate that this region of integrase is involved primarily in protein-protein interactions and/or recognition of the viral DNA substrate.

The central core. The mutations D116N and E152Q, which introduced amide groups at these two highly conserved positions, abolished all detectable in vitro activities. Similar results have recently been reported for the 3′ processing and DNA strand transfer activities of mutant HIV-1 integrase proteins containing substitutions at these positions (11, 31). Although we cannot exclude the possibility that these amino acid changes have long-range effects on protein structure, the loss of all function upon mutation of these residues suggests that they may form part of the active site. Our experiments do not address the role of the C-terminal part of integrase protein, but the divergence of the amino acid sequence of this region between different retroviruses suggests that it is not a structurally conserved part of the protein, as would be expected for the active site.

Participation of negatively charged amino acid side chains in the active site might be anticipated from their known role in coordinating divalent metal ions in enzymatic catalysis of several other phosphoryl transfer reactions. For example, a pair of metal ions, positioned by interactions with nearby negatively charged amino acid side chains, is involved in the 3′-5′ exonuclease reaction catalyzed by DNA polymerase I (2). The pair of metal ions is thought to be involved in both formation of the attacking hydroxyl group and the stabilization and resolution of the reaction intermediate. A similar role for the conserved acidic residues has been suggested to be involved in catalysis by integrase (31). Tentative support for the notion that D-116 lies within the active site comes from the relatively greater abundance of cyclic dinucleotide cleavage product made in reactions with N117Q. Lengthening the side chain of this adjacent amino acid residue presumably modifies the local environment so as to increase the relative efficiency of nucleophilic attack by the 3′ end of the viral DNA rather than water.

How are the activities of integrase protein related to each other? Our results are consistent with a common active site for 3′ processing, DNA strand transfer, and disintegration,

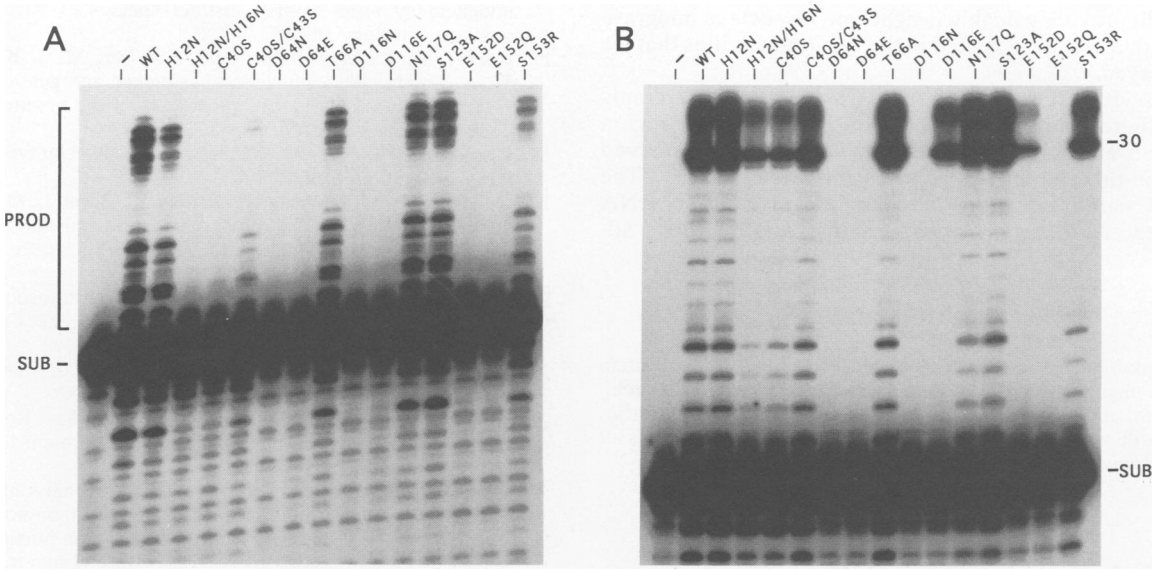

FIG. 6. DNA strand transfer and disintegration reactions of mutant integrase proteins. (A) DNA strand transfer activity of the proteins was assayed at a protein concentration of 0.30 μM. Reactions were analyzed on a denaturing polyacrylamide gel. Integrase was omitted from the reaction in the lane marked −; the lane marked WT contained a reaction with wild-type protein; the other lanes contained reactions with the indicated mutant proteins. Other labeling is the same as in Fig. 4B. (B) Disintegration activity of the same proteins was tested at a protein concentration of 0.30 μM. Reaction products were analyzed on a denaturing polyacrylamide gel. Integrase was omitted from the reaction in the lane marked −; the lane marked WT contained a reaction with wild-type protein; the other lanes contained reactions with the noted mutant proteins. Other labeling is the same as described in the legend to Fig. 4C.

as previously suggested by the results of the stereochemical analysis of the 3′ processing and DNA strand transfer reactions (13). However, these reactions may differ in the protein-protein and protein-DNA interactions involved. Mutations that discriminate between these activities may provide information on the parts of the protein that are involved in such interactions. All of the proteins that we tested with mutations in the His-Cys region were quite active in the disintegration assay but impaired to a greater extent in 3′ processing and DNA strand transfer activities. These data suggest that some interactions involving this region that are important for cleavage and strand transfer are dispensable for disintegration. Protein-protein interactions and recognition of the viral DNA end are obvious candidates. We also find that whereas the efficiency of the 3′ processing and DNA strand transfer reactions show a distinctly nonlinear dependence on integrase concentration, with little product generated below a threshold concentration, disintegration can occur at lower protein concentrations (Fig. 4). One possible interpretation is that a higher multimeric state of integrase is required for 3′ processing and DNA strand transfer than for disintegration.

We note that the in vitro reactions that we have used to assay the effects of mutations probably do not test the ability of integrase to perform all of the functions that are required of it in vivo. For example, a majority of the in vitro DNA strand transfer products result from insertion of a single processed viral DNA end substrate molecule into a single strand of the target DNA, rather than concerted insertion of a pair of viral DNA ends. Coordinated insertion of pairs of ends may require a higher oligomeric state of integrase that is not efficiently formed in our reactions. These assays may not detect the effects of mutation of residues that are normally involved in such oligomerization or interactions between integrase and other viral proteins that may be important in vivo. It would be of particular interest to assay the replication of HIV-1 containing the H12N mutation in vivo, since this mutant is quite active in all three in vitro assays. Failure of virus with this mutation to replicate in vivo might provide valuable insights into aspects of integrase function that are not reflected in the in vitro activities that we have assayed.

The in vitro activities of the HIV-1 integrase mutants reported here points to a key role for the central core region in catalysis and suggests that the His-Cys region is involved in protein-protein and/or protein-DNA interactions. The nature of such interactions between integrase and its DNA substrates are intriguing questions that are yet to be addressed.

ACKNOWLEDGMENTS

We thank Moshe Sadofsky for assistance in preparing the protein sequence alignment and Frederic Bushman, Kiyoshi Mizuuchi, David Roth, and M. Sadofsky for careful reading of the manuscript.

This work was supported in part by the National Institutes of Health Intramural AIDS Targeted Antiviral Program.

REFERENCES

1. **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 nonhuman cells transfected with an infectious molecular clone. J. Virol. **59**:284–291.
2. **Beese, L. S., and T. A. Steitz.** 1991. Structural basis for the 3′-5′ exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. EMBO J. **10**:25–33.
3. **Brown, P. O.** 1990. Integration of retroviral DNA. Curr. Top. Microbiol. Immunol. **157**:19–48.
4. **Brown, W. E., and F. Wold.** 1973. Alkyl isocyanates as active-site-specific reagents for serine proteases. Biochemistry **12**:828–834.
5. **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.
6. **Bushman, F. D., T. Fujiwara, and R. Craigie.** 1990. Retroviral DNA integration directed by HIV integration protein in vitro. Science **249**:1555–1558.
7. **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.
8. **Copeland, T. D., G. F. Gerard, C. W. Hixson, and S. Oroszlan.** 1985. Amino- and carboxyl-terminal sequence of Moloney murine leukemia virus reverse transcriptase. Virology **143**:676–679.
9. **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.
10. **Donahue, P. R., E. A. Hoover, G. A. Beltz, N. Riedel, V. M. Hirsch, J. Overbaugh, and J. I. Mullins.** 1988. Strong sequence conservation among horizontally transmissible, minimally pathogenic feline leukemia viruses. J. Virol. **62**:722–731.
11. **Drelich, M., 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.
12. **Engelman, A., and R. Craigie.** Unpublished data.
13. **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.
14. **Fayet, O., P. Ramond, P. Polard, M. F. Prere, and M. Chandler.** 1990. Functional similarities between the IS*3* family of bacterial insertion elements? Mol. Microbiol. **4**:1771–1777.
15. **Flugel, R. M., A. Rethwilm, B. Maurer, and G. Darai.** 1987. Nucleotide sequence analysis of the *env* gene and its flanking regions of the human spumaretrovirus reveals two novel genes. EMBO J. **6**:2077–2084.
16. **Fomsgaard, A., V. M. Hirsch, J. S. Allan, and P. R. Johnson.** 1991. A highly divergent proviral DNA clone of simian immunodeficiency virus from a distinct species of African green monkey. Virology **182**:397–402.
17. **Garvey, K. J., M. S. Oberste, J. E. Elser, M. J. Braun, and M. A. Gonda.** 1990. Nucleotide sequence and genome organization of biologically active proviruses of the bovine immunodeficiency-like virus. Virology **175**:391–409.
18. **Grandgenett, D. P., and S. R. Mumm.** 1990. Unraveling retrovirus integration. Cell **60**:3–4.
19. **Guyader, M., M. Emerman, P. Sonigo, F. Clavel, L. Montagnier, and M. Alizon.** 1987. Genome organization and transcription of the human immunodeficiency virus type 2. Nature (London) **326**:662–669.
20. **Heidmann, O., and T. Heidmann.** 1991. Retrotransposition of a mouse IAP sequence tagged with an indicator gene. Cell **64**:159–170.
21. **Herr, W.** 1984. Nucleotide sequence of AKV murine leukemia virus. J. Virol. **49**:471–478.
22. **Hippenmeyer, P. J., and D. P. Grandgenett.** 1984. Requirement of the avian retrovirus pp32 DNA binding protein for replication. Virology **137**:358–370.
23. **Hirsch, V. M., R. A. Olmstead, M. Murphey-Corb, R. H. Purcell, and P. R. Johnson.** 1989. Simian immunodeficiency virus from Sooty mangabeys: an African non-human primate lentivirus closely related to HIV-2. Nature (London) **339**:389–392.
24. **Hizi, A., and S. H. Hughes.** 1988. Expression of the Moloney murine leukemia virus and human immunodeficiency virus integration proteins in *Escherichia coli*. Virology **167**:634–638.
25. **Houmard, T., and G. R. Drapeau.** 1972. Staphylococcal protease: a proteolytic enzyme specific for glutamoyl bonds. Proc.

Natl. Acad. Sci. USA 69:3506–3509.
26. **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.
27. **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.
28. **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.
29. **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.
30. **Kornfeld, H., N. Riedel, G. A. Viglianti, V. Hirsch, and J. I. Mullins.** 1987. Cloning of HTLV-4 and its relation to simian and human immunodeficiency viruses. Nature (London) 326:610–613.
31. **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.
32. **Lafemina, R. L., P. L. Callahan, and M. G. Corgingley.** 1991. Substrate specificity of recombinant human immunodeficiency virus integrase protein. J. Virol. 65:5624–5630.
33. **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.
34. **Lightfoote, M. M., J. E. Coligan, T. M. Folks, A. S. Fauci, M. A. Martin, and S. Venkatesan.** 1986. Structural characterization of reverse transcriptase and endonuclease polypeptides of the acquired immunodeficiency syndrome virus. J. Virol. 60:771–775.
35. **Malik, K. T. A., J. Even, and A. Karpas.** 1988. Molecular cloning and complete nucleotide sequence of an adult T cell leukemia virus type I (ATLV/HTLV-I) isolate of Caribbean origin: relationship to other members of the ATLV/HTLV-I subgroup. J. Gen. Virol. 69:1695–1710.
36. **Matsudaira, P.** 1987. Sequence from picomole quantities electroblotted onto polyvinylidene difluoride membranes. J. Biol. Chem. 262:10035–10038.
37. **Miller, A. D., and I. M. Verma.** 1984. Two base changes restore infectivity to a noninfectious molecular clone of Moloney murine leukemia virus. J. Virol. 49:214–222.
38. **Moore, R., M. Dixon, R. Smith, G. Peters, and C. Dickson.** 1987. Complete nucleotide sequence of a milk-transmitted mouse mammary tumor virus: two frameshift suppression events are required for translation of *gag* and *pol*. J. Virol. 61:480–490.
39. **Ratner, L., A. Fisher, L. L. Jagodzinski, H. Mitsuya, R.-S. Liou, R. C. Gallo, and F. Wong-Staal.** 1987. Complete nucleotide sequences of functional clones of the AIDS virus. AIDS Res. Hum. Retroviruses 3:57–69.
40. **Rethwilm, A., G. Baunach, K.-O. Netzer, B. Maurer, B. Borisch, and V. ter Meulen.** 1990. Infectious DNA of the human spumaretrovirus. Nucleic Acids Res. 18:733–738.
41. **Rice, N. R., R. M. Stephens, A. Burney, and R. V. Gilden.** 1985. The *gag* and *pol* genes of bovine leukemia virus: nucleotide sequence and analysis. Virology 142:357–377.
42. **Rosenberg, A. H., B. N. Lade, D. Chui, S.-W. Lin, J. J. Dunn, and F. W. Studier.** 1987. Vectors for selective expression of cloned DNAs by T7 RNA polymerase. Gene 56:125–135.
43. **Rowland, S.-J., and K. G. H. Dyke.** 1990. Tn552, a novel transposable element from *Staphylococcus aureus*. Mol. Microbiol. 4:961–975.
44. **Saltarelli, M., G. Querat, D. A. Konings, R. Vigne, and J. E. Clements.** 1990. Nucleotide sequence and transcriptional analysis of molecular clones of caprine arthritis encephalitis virus which generate infectious virus. Virology 179:347–364.
45. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463–5467.
46. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell 32:853–869.
47. **Schwartz, R. M., and M. O. Dayhoff.** 1978. Matrices for detecting distant relationships, p. 353–358. *In* M. O. Dayhoff (ed.), Atlas of protein sequence and structure, vol. 5, suppl. 3. National Biochemical Research Foundation, Washington, D.C.
48. **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.
49. **Shimotohno, K., Y. Takahashi, N. Shimizu, T. Gojobori, D. W. Golde, I. S. Y. Chen, M. Miwa, and T. Sugimura.** 1985. Complete nucleotide sequence of an infectious clone of human T-cell leukemia virus type II: an open reading frame for the protease gene. Proc. Natl. Acad. Sci. USA 82:3101–3105.
50. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukemia virus. Nature (London) 293:543–548.
51. **Sonigo, P., C. Barker, E. Hunter, and S. Wain-Hobson.** 1986. Nucleotide sequence of Mason-Pfizer monkey virus: an immunosuppressive D-type retrovirus. Cell 45:375–385.
52. **Stephens, R. M., J. W. Casey, and N. R. Rice.** 1986. Equine infectious anemia virus *gag* and *pol* genes: relatedness to visna and AIDS viruses. Science 231:589–594.
53. **Studier, F. W., and B. A. Moffatt.** 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. 189:113–130.
54. **Talbott, R. L., E. E. Sparger, K. M. Lovelace, W. M. Fitch, N. C. Pedersen, P. A. Luciw, and J. H. Elder.** 1989. Nucleotide sequence and genomic organization of feline immunodeficiency virus. Proc. Natl. Acad. Sci. USA 86:5743–5747.
55. **van Gent, D. C., Y. Elgersma, M. W. J. Bolk, C. Vink, and R. H. A. Plasterk.** 1991. DNA binding of the integrase proteins of human immunodeficiency viruses types 1 and 2. Nucleic Acids Res. 19:3821–3827.
56. **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.
57. **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.
58. **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.
59. **Whetter, L., D. Archambault, S. Perry, A. Gazit, L. Coggins, A. Yaniv, D. Clabough, J. Dahlberg, and S. Tronick.** 1990. Equine infectious anemia virus derived from a molecular clone persistently infects horses. J. Virol. 64:5750–5756.
60. **Wilhelmsen, K. C., K. Eggleton, and H. Temin.** 1984. Nucleic acid sequences of the oncogene v-*rel* in reticuloendotheliosis virus strain T and its cellular homolog, the proto-oncogene c-*rel*. J. Virol. 52:172–182.
61. **Ymer, S., W. Q. J. Tucker, H. D. Campbell, and I. G. Young.** 1986. Nucleotide sequence of the intracisternal A-particle genome inserted 5′ to the interleukin-3 gene of the leukaemia cell line WEHI-3B. Nucleic Acids Res. 14:5901–5918.