Comparative studies of bacterially expressed integrase proteins of caprine arthritis–encephalitis virus, maedi–visna virus and human immunodeficiency virus type 1

Klaus D. Störmann, Michaela C. Schlecht and Eberhard Pfaff*

Federal Research Centre for Virus Diseases of Animals, Paul-Ehrlich-Strasse 28, PO Box 1149, 72001 Tübingen, Federal Republic of Germany

Integrase (IN) proteins mediate an essential step in retroviral life cycles, the integration of reverse-transcribed viral DNA into the host genome. To create tools for direct comparative investigations, hexahistidine-tagged IN proteins of the phylogenetically related lentiviruses caprine arthritis–encephalitis virus (CAEV), maedi–visna virus (MVV) and human immunodeficiency virus type 1 (HIV-1) were expressed in Escherichia coli. After purification by affinity chromatography, the active enzymes were compared in vitro for their site-specific cleavage, integration and disintegration activities on cognate and non-cognate oligonucleotide substrates. It was found that CAEV IN and MVV IN catalyse both site-specific cleavage and disintegration with high efficiencies, reduced substrate specificities and similar reaction patterns. Comparisons with the respective activities of HIV-1 IN revealed basic functional similarities as well as considerable differences such as more restricted substrate requirements for site-specific cleavage. On the other hand, all three enzymes catalyse disintegration almost independent of the substrate origin. Furthermore, MVV IN was shown to join oligonucleotides as efficiently as HIV-1 IN, albeit with reduced substrate specificity. In contrast, no detectable strand transfer activities occurred with CAEV IN.

Introduction

Caprine arthritis–encephalitis virus (CAEV) and maedi–visna virus (MVV) are closely related lentiviruses infecting goat and sheep. Unlike human immunodeficiency virus (HIV), the ovine–caprine lentiviruses do not cause immunodeficiencies, but interstitial pneumonia,encephalomyelitis, mastitis, arthritis and generalized wasting (for reviews, see Cheevers & McGuire, 1988; Perk, 1988; Narayan et al., 1993). Each of these diseases is characterized by the long incubation period and the chronic, slowly progressive course typical of lentiviral infections. Since CAEV and MVV share numerous biological similarities with HIV, comparative studies of their enzymes have both scientific and clinical relevance. In particular, integrase (IN) proteins have been recognized as important targets for antilentiviral strategies.

Integration of a DNA copy of genomic viral RNA into the host cell DNA is essential for retroviral replication (LaFemina et al., 1992; Sakai et al., 1993). The crucial steps of this process are catalysed by IN, which is the only viral protein required in vitro (Katzman et al., 1989; Craigie et al., 1990; Katz et al., 1990). After infection of a permissive cell, reverse transcriptase converts the retroviral RNA into a double-stranded, blunt-ended DNA molecule. Prior to integration, IN removes the terminal two bases 3' of a highly conserved CA dinucleotide (referred to as site-specific cleavage, endonucleolytic cleavage or 3' processing) (Katzman et al., 1989; Craigie et al., 1990; Bushman & Craige, 1991; Engelman et al., 1991). The new OH groups generated at both ends of the linear viral cDNA are used as nucleophiles to break phosphodiester bonds in the target DNA and to establish new covalent bonds between viral and host DNA in a one-step transesterification reaction (referred to as strand transfer, integration or DNA joining) (Bushman et al., 1990; Craigie et al., 1990; Bushman & Craigie, 1991; Engelman et al., 1991). The broken phosphodiester bonds are 4 to 6 bp apart from each other. Although target site selection by IN was considered random, preferentially used sequence and chromatin structures have been observed recently (Craigie, 1992; Leavitt et al., 1992; Hong et al., 1993).

In the integration intermediate resulting from strand transfer the two non-paired nucleotides at the 5' ends of the viral DNA remain unjoined and are presumably removed by the cellular repair machinery. The single-stranded gaps at both ends of the integrated retrovirus
are probably also filled in this step. Finally, the resulting proviral DNA has lost 2 bp from each end and is flanked by short, direct repeats of the target DNA (for recent reviews, see Brown, 1990; Goff, 1992; Mizuuchi, 1992; Vink & Plasterk, 1993).

IN proteins also catalyse disintegration, an in vitro reversal of strand joining reactions. In the disintegration process a complex branched substrate, resembling an integration intermediate resulting from strand transfer, is cleaved into two molecules corresponding to the viral and host DNAs (Chow et al., 1992). Whether disintegration plays a role in vivo still remains unclear. Like disintegration, site-specific cleavage and strand transfer reactions can be investigated in vitro using double-stranded oligonucleotide substrates representing the outer ends of viral long terminal repeats (LTRs) (Katzman et al., 1989).

Previous studies demonstrated that recombinant integrases possess both the endonucleolytic cleavage and the strand transfer activities necessary for integration of viral DNA (Katzman et al., 1989; Bushman et al., 1990; Craigie et al., 1990; Bushman & Craigie, 1991; Pahl & Flügel, 1993; Bushman & Wang, 1994; Vink et al., 1994). Comparisons of amino acid sequences of the integrases of numerous retroviruses and retrotransposons and analyses of mutant HIV-1, HIV-2 and Rous sarcoma virus (RSV) IN proteins have revealed three different regions that are responsible for the enzymatic functions.

The N-terminal zinc-finger-like part of the enzyme, designated the HHCC region, was shown to bind Zn$^{2+}$ (Burke et al., 1992; McCuen et al., 1992; Bushman et al., 1993). This motif is well conserved among retroviruses and may function in the recognition of the viral DNA ends (Van Gent et al., 1992; Vincent et al., 1993). Mutations within the HHCC region lead to defects in $3'$ processing and strand transfer. The second conserved motif is defined by three amino acid residues in a most highly conserved spatial arrangement. This ‘D,D(35)E’ motif is located in the central region of the protein and is also present in the transposases of several bacterial transposons (Khan et al., 1991; Kulkosky et al., 1992). Mutations in any one of these residues completely abolish all IN activities (Drelich et al., 1992; Engelman & Craigie, 1992; Van Gent et al., 1992). Therefore, it is thought that the two invariant Asp residues and the Glu residue form the active centre of IN responsible for all polynucleotidyl transfer reactions (Bushman et al., 1993; Vink et al., 1993). The carboxy terminus is the least conserved region among retroviral integrases. This domain was shown to bind both the viral DNA ends and non-specific DNA with similar affinity (Vink et al., 1991a; LaFemina et al., 1992; Schauer & Billich, 1992; Woerner et al., 1992). How the IN protein discriminates between viral DNA ends and other DNA is not yet known. In addition, it was indicated recently that IN is active as an oligomer (Engelman et al., 1993; Kalpana & Goff, 1993; Van Gent et al., 1993).

Although the IN proteins of several oncoviruses and HIV have been the subjects of intensive studies, little is known about the integrases of animal lentiviruses. In the present study, we have therefore expressed, purified and tested the IN proteins of CAEV and MVV. Furthermore, all enzyme functions were directly compared with those of HIV-1 IN which was expressed and purified under identical conditions.

### Methods

**Viruses, cells and cell culture.** Monolayers of primary goat synovial membrane (GSM) cells and primary fetal lamb testes (FLT) cells were used for propagation of American strain 75-G63 of CAEV (ATCC, Rockville, Md., USA) and German strain 461 of MVV (BFAV, Tübingen, Germany), respectively. Cells were grown at 37 °C in a humidified atmosphere of 5% CO$_2$ in minimal essential medium (MEM) supplemented with 1 mM-glutamine and 15% fetal calf serum (for expansion of uninfected cells) or 2% lamb serum (for virus propagation).

**Infection of cells and preparation of low-molecular-mass DNA.** Confluent GSM and FLT cells were infected with supernatants from cultures which had completely lysed after preceding infection with CAEV 75-G63 and MVV 461, respectively; 48 h after infection the low-molecular-mass DNA was extracted according to the method of Hirt (1967).

**Amplification of IN coding DNA.** The termini of CAEV IN and MVV IN were identified comparing the predicted Gag–Pol precursors of both viruses with other retroviral IN proteins. PCR amplification of IN coding regions was performed using Hirt DNA from infected cells as template and oligonucleotides introducing unique Neol or BglII restriction sites (in bold type) as primers. Oligonucleotide primers were synthesized on a Millipore Biosearch 8700 DNA synthesizer. The oligonucleotides used for the amplification of CAEV IN coding DNA were CAEV IM1+ (5' GGATTGTTCCACAGCCATGGTTTGGGATAGAAAATATCCCTTTGGCAGA 3'; nucleotides encoding the N-terminal Trp residue of CAEV IN are underlined) and CAEV IM2+ (5' ATGGTAATTCTGGTCCCTGTTCCCTAGATCTTTCTCCTTTGCTGGTGGGC 3'; the underlined nucleotides are complementary to the last codon of CAEV IN; except that a 'T' was replaced by an 'A' to create a BglII restriction site). Oligonucleotides used for the amplification of MVV IN coding DNA were MVVM1+ (5' ATCAAGGGTTGGATCGACAGCCATGGTGATTGAAAACACCTTCA 3'; nucleotides encoding the N-terminal Trp residue of MVV IN are underlined) and MVVM1-2+ (5' GGTCAATTCATTCCCAATGTTAGATCTTTGATATTCTTTGTCGGTGGA 3'; the underlined nucleotides correspond to the last codon of MVV IN, except that the authentic sequence ‘GGC’ was changed to ‘ATC’ to create a BglII restriction site). 

HIV-1 IN coding DNA was derived from plasmid pNL4-3 containing a full-length clone of HIV-1 obtained through recombinant of strains NY5 and BRU (Adachi et al., 1986). The oligonucleotide primers which also introduced Neol and BglII restriction sites used for amplification of HIV-1 IN were HIVIM1+ (5' TGCTACGTCGGAGAATCCAGA- CCATGGATTATTTAGATGGAATGAA 3'; nucleotides encoding the N-terminal Phe residue of HIV-1 IN, positions 4230 to 4232 of the HIV-1 NL4-3 sequence, are underlined) and HIVIM2+ (5' ACTATCTTCTGTCATTCACTCCTACATTGTGACCA 3'; the underlined nucleotides are complementary...
to the last codon of HIV-1 IN, positions 5091 to 5093 of the HIV-1 NL4-3 sequence). PCR was performed in a 100 μl mixture containing 10 mM-Tris-HCl, pH 8.0, 50 mM-KCl, 1.5 mM-MgCl₂, 0.01 % gelatine, 0.2 mM of each dATP, dCTP, dGTP and dTTP, 50 pmol of each primer, 1 μg of Hirt DNA or 100 ng of pNL4-3 DNA, and 2.5 U Taq DNA polymerase (Stratagene). The conditions used for 30 cycles of denaturating polyacrylamide gels. The sequences of the 20-mers in its first 6 bp, whereas the remaining sequence was chosen to generate a ‘dumbbell’ structure (Fig. 1b; Vincent et al., 1993). A schematic expression and native purification of IN proteins by Ni²⁺ chelate affinity chromatography were carried out as described by Drelich et al. (1992). The selective removal of contaminants and the elution of hexahistidine-tagged IN proteins were achieved through competition with increasing imidazole concentrations. Afterwards, the imidazole was removed by dialysis against storage buffer (50 mM-Tris-HCl, pH 7.6, 20 mM-β-mercaptoethanol, 0.1 mM-EDTA, pH 8.0, 25%, v/v, glycerol and 500 mM-NaCl). Determination of protein concentrations was performed according to the method of Bradford (1976) using bovine gamma globulin as a standard (Bio-Rad).

**Tests for endonucleolytic cleavage, integration and disintegration activities.** The 12 substrates used to test for IN activities are listed in Fig. 1. The different oligonucleotides, 20 and 43 bases in length, were synthesized as described above and purified by electrophoresis on denaturing polyacrylamide gels. The sequences of the 20-mers correspond to the outer U5 and U3 LTR ends of viral DNA of CAEV and MVV U5 (positions 75-G63, MVV 461 and HIV-1 NL4-3 (Fig. 1a). The 43-mers also correspond to the U5 and U3 LTR ends of these virus isolates but only in its first 6 bp, whereas the remaining sequence was chosen to generate a ‘dumbbell’ structure (Fig. 1b; Vincent et al., 1993). A schematic

![Diagram](Image)

**Fig. 1.** Oligonucleotide substrates used to test for IN activities. (a) Sequence and structure of double-stranded substrates for cleavage and integration assays. U5 and U3 substrates represent the respective viral DNA ends of the indicated virus. (b) Sequence and dumbbell structure of substrates for disintegration assays. Excerpt for the four T nucleotides, the branches correspond to the respective viral DNA ends of the indicated viruses. The parts with an identical base composition represent an arbitrary host sequence. Since the last nine base pairs at the 3' ends of CAEV and MVV U5 are identical, CAEV and MVV U5 DBS have the same sequences. The position of the radiolabel is represented by ³²P. Predominant cleavage sites are marked by arrows. The conserved dinucleotide CA is in bold type.
For site-specific cleavage and integration reactions the 5' termini of plus strands were end-labelled using T4 polynucleotide kinase (Biolabs) and \([p^{32}\text{P}]\text{ATP} (5000 \text{ Ci/mmole, Amersham}) and annealed with their complementary strands in 10 mM-Tris-HCl, pH 7.6 and 150 mM NaCl. For disintegration reactions the dumbbell oligonucleotides were labelled as described above and self-annealed. One picomole of the different substrates was incubated with 1 ng (~ 5 pmol) of purified IN. The 10 \(\mu\)l standard reaction mixtures further contained 25 mM-Tris-HCl, pH 8.0, 1 mM-DTT, either 2 mM-Mg\(^{2+}\) or 2 mM-Mn\(^{2+}\), and the components of the protein and oligonucleotide buffers in one-tenth of their given concentrations. Therefore, the final content of NaCl in the assays was 65 mM. After 90 min at 37 °C the reactions were stopped by addition of 7 \(\mu\)l of dye containing formamide (95% formamide, 20 mM-EDTA, 0.05% bromphenol blue and 0.05% xylene cyanol). Samples (3 \(\mu\)l) were heated to 95 °C for 5 min, electrophoresed on 15% denaturing polyacylamide gels, and autoradiographed. Afterwards, in the case of disintegration assays, the major products were cut-out, eluted, and characterized by Maxam–Gilbert sequencing (Maxam & Gilbert, 1977).

**Results**

**Cloning, expression and purification of CAEV IN, MVV IN and HIV-1 IN**

The IN genes from CAEV 75-G63, MVV 461 and HIV-1 NL4-3 were amplified by PCR and subcloned into the expression vector pQE-60. For bacterial expression of CAEV IN, an authentic ATG codon was used which is located two amino acids upstream from the first N-terminal amino acid of the enzyme. In the case of MVV IN and HIV-1 IN, artificial ATG codons were introduced at corresponding positions. Together with six histidine codons indispensable for purification using Ni-NTA, new stop codons were added at the 3' ends of the genes. To make sure that all components of the constructs were in-frame and that no PCR mediated mutations had occurred, the coding regions of the resulting plasmids were sequenced completely.

A comparison of the deduced amino acid sequences of CAEV IN, MVV IN and HIV-1 IN is shown in Fig. 3. Identity ranges from 75% between CAEV IN and MVV IN, 33% between CAEV IN and HIV-1 IN, and 31% between MVV IN and HIV-1 IN. Moreover, the IN proteins of CAEV 75-G63 and MVV 461 differ not only from each other, but also from integrases of the respective prototype isolates CAEV CO (Saltarelli et al., 1990) and MVV 1514 (Braun et al., 1987). In particular, the identity between IN proteins of the American strains 75-G63 and CO of CAEV is 91%, whereas the identity between IN proteins of the German strain 461 and the Icelandic strain 1514 of MVV reaches only 86%. Therefore, it was expected that the properties of MVV 461 IN could differ slightly from those of MVV 1514 IN.
Comparative studies of lentiviral IN proteins

Fig. 3. Alignment of the amino acid sequences of CAEV 75-G63, MVV 461 and HIV-1 NL4-3 IN proteins. Amino acid residues which are highly conserved within the lentivirus group are boxed. Residues also conserved in integrases and transposases of oncoviruses, foamy viruses, retrotransposons and bacterial IS elements are indicated by additional shading. The similarity of respective amino acid residues is shown by lines, colons, points and blanks. The comparison is based on computational analysis using the UWGCG software (Devereux et al., 1984).

which was shown very recently to be a fully functional retroviral integrase (Katzman & Sudol, 1994).

Like HIV-1 IN, the N-terminal halves of CAEV IN and MVV IN contain seven residues highly conserved among retroviruses, constituting the Zn$^{2+}$ binding HHCC motif and the D,D(35)E motif. In addition, 40 residues are found which are conserved only among lentiviruses. The conservation of these residues was determined by comparing more than 40 predicted lentiviral IN amino acid sequences (data not shown). The integrases of CAEV and MVV each contain two additional amino acids at the C-terminal ends of their HHCC regions but are ten and nine amino acids shorter at their respective C termini.

Recombinant integrase expression vectors were introduced into E. coli and after induction large amounts of all three enzymes were synthesized. As expected from the calculated values, IN proteins with molecular masses of approximately 33 kDa are found in the cases of CAEV IN, MVV IN and HIV-1 IN, respectively (Fig. 4, lanes 6, 10 and 14). Expression of the latter results in an additional product of apparently 19 kDa (Fig. 4, lane 14). N-terminal sequencing revealed that this protein was translated using an internal Shine–Dalgarno sequence upstream from Met$^{144}$ of the HIV-1 IN amino acid sequence. In contrast to the observed chromatographic behaviour, its calculated molecular mass is only 16.2 kDa.

The recombinant IN proteins were purified from lysed bacteria by extraction with high salt, followed by Ni-NTA affinity chromatography. This procedure resulted in milligram quantities of soluble and enzymatically active IN proteins which were judged to be more than 90% pure based on analysis of Coomassie brilliant blue-stained gels. The results of typical integrase purifications are shown in Fig. 4, lanes 8, 12 and 16. The truncated HIV-1 IN protein does not elute under the conditions used to wash and recover full length HIV-1
IN and is therefore removed by chromatography. For this reason no disturbing influences on the tests for IN activity were expected.

To obtain negative controls for enzyme activity assays, crude extracts from E. coli M15 transformed with the unmodified expression vector pQE-60 were purified as described above (Fig. 4, lanes 1 to 4). Aliquots of eluates from this purification procedure were used to substitute for the IN proteins in control reactions.

**Endonuclease activities of CAEV IN, MVV IN and HIV-1 IN**

Site-specific cleavage activities of the three different IN proteins were investigated using radiolabelled oligonucleotides (Fig. 2a). The double-stranded 20-mer substrates, identical in sequence to the viral DNA ends of CAEV 75-G63, MVV 461 and HIV-1 NL4-3, are shown in detail in Fig. 1(a). Like HIV-1 IN, CAEV IN and MVV IN were expected to specifically remove two nucleotides from the 3' ends of their 5'-end-labelled strands.

CAEV IN and MVV IN were tested and found to be active over the same range of reaction conditions as described previously for HIV-1 IN (Drelich et al., 1992; Vincent et al., 1993). Therefore, one standard condition was chosen which is close to the optimum for each of the three enzymes (25 mM-Tris-HCl, pH 8.0, 1 mM-DTT, 2 mM-Mn²⁺, 90 min at 37 °C). Since Mg²⁺ is the predominant divalent cation in vivo, activity assays were also performed with MgCl₂ instead of MnCl₂.

In the presence of Mg²⁺ the CAEV IN and MVV IN proteins demonstrate weak but detectable 3' processing on both their cognate U5 (removal of G and C) and U3 (removal of G and T) substrates (data not shown). Using the same divalent cation, HIV-1 IN cleaves its U3 substrate with a comparable efficiency, whereas a significantly higher activity is found on its U5 substrate (data not shown).

More efficient 3' processing occurs in the presence of
Comparative studies of lentiviral IN proteins

Mn\(^{2+}\) (Fig. 5). CAEV IN and MVV IN cleave their respective U5 and U3 substrates with very similar productivities resulting in specifically shortened 18-mer \((n-2)\) products. HIV-1 IN demonstrates a comparable activity with its U5 substrate, whereas its U3 substrate is less efficiently cleaved (Fig. 5a, b, lanes 12). Besides the 18-mer, an additional product of 19 nucleotides \((n-1)\) appears in both reactions with HIV-1 IN. Respective products were also found in the presence of Mg\(^{2+}\). Although the control lanes do not contain any detectable cleavage activities (Fig. 5a, b, lanes 12), the 19-mers could be due to a low level of co-purified \(E. coli\) nucleases. Therefore, different HIV-1 IN preparations were further purified under more stringent washing conditions or cation-exchange chromatography. However, all preparations containing active HIV-1 IN also produced the 19-mer.

### Substrate specificities of endonuclease activities

To determine the specificity of site-specific cleavage, the IN proteins of CAEV, MVV and HIV-1 were also tested on heterologous U5 and U3 substrates. As shown in Fig. 5(a), CAEV IN and MVV IN cleave the CAEV U5 substrate with very similar efficiencies (lanes 2 and 3). Since the HIV-1 IN preparation possibly contains a contaminating nuclease activity, it is difficult to distinguish an integrase-dependent \(n-2\) endonuclease activity from two sequential nuclease-dependent \(n-1\) cuts. Therefore, if at all, HIV-1 IN cleaves CAEV U5 rather poorly and only the 19-mer is found in a high amount (Fig. 5a, lane 4). The same holds for HIV-1 IN endonuclease reactions on MVV U5 (Fig. 5a, lane 8), CAEV U3 (Fig. 5b, lane 4) and MVV U3 (Fig. 5b, lane 8). With the same substrates CAEV IN and MVV IN demonstrate clearly visible cleavage activities, but CAEV IN converts them less efficiently (Fig. 5a, lanes 6 and 7; Fig. 5b, lanes 2, 3, 6 and 7).

In addition, HIV-1 U5 and U3 substrates were investigated with the different IN proteins (Fig. 5a, b, lanes 9 to 12). All three integrases cleave the respective substrate with comparable efficiencies, although differences exist between HIV-1 U5 and HIV-1 U3. In contrast to CAEV and MVV U5 and U3 which are cleaved with nearly the same efficiencies, HIV-1 U5 is a better substrate than HIV-1 U3.
Strand transfer activities of CAEV IN, MVV IN and HIV-1 IN

A further enzymatic activity associated with IN proteins is the ability to insert viral DNA into host DNA. This process can be mimicked using the same double-stranded substrates as for endonucleolytic cleavage (Fig. 2a, b). The positive result of such an in vitro integration assay is a ladder of bands representing elongated oligonucleotides. Favourable reaction conditions for both endonucleolytic cleavage and strand transfer were found to be similar in the cases of MVV IN and HIV-1 IN.

Elongated strand transfer products situated above the initial 20-mer band were detected after prolonged exposure times of the autoradiographs presented in Fig. 5(a, b). On the other hand, no integration products occurred in reactions carried out with Mg~2+, not even after extended exposure (data not shown). Thus, the in vitro integration activities of the three purified IN proteins seem to be extremely weak or completely missing if Mg~2+ is the divalent cation.

In the presence of Mn~2+, HIV-1 IN integrates its U5 substrate with high efficiency and also, albeit with considerably reduced efficiency, its U3 substrate (Fig. 6a, b, lanes 12). In contrast, MVV IN integrates its U3 substrate more efficiently than its U5 substrate (Fig. 6a, b, lanes 7). In the case of CAEV IN, no strand transfer products are detectable after incubation with the cognate U5 or U3 substrates (Fig. 6a, b, lanes 2). Since it was thought that this finding could result from artefactual misfolding or more complex reaction requirements, different enzyme preparations and reaction mixtures were tested. Unfortunately, all efforts to obtain a CAEV IN protein active in strand transfer failed.

Substrate specificities of strand transfer activities

As described for endonucleolytic cleavage, the substrate specificities of IN proteins were also investigated for integration. Again, CAEV IN shows no detectable strand transfer activities, whereas very efficient DNA joining is seen with MVV IN tested on CAEV and MVV U3 oligonucleotides (Fig. 6b, lanes 3 and 7). Moreover, MVV IN joins molecules of the HIV-1 U5 substrate (Fig. 6a, lane 11), the MVV U5 and HIV-1 U3 substrates (Fig. 6a, lane 7; Fig. 6b, lane 11) and the CAEV U5 substrate (Fig. 6a, lane 3), although with lower efficiencies. With respect to the HIV-1 U3 substrate, the efficiency of integration seems to be even slightly higher using MVV IN instead of HIV-1 IN (Fig. 6b, lanes 11 and 12). A more efficient integration with the HIV-1 IN protein is only found in the case of HIV-1 U5 (Fig. 6a, lane 12) but no strand transfer products occur with non-cognate substrates (Fig. 6a, b, lanes 4 and 8).

Since the assay did not use ‘pre-cleaved’ substrates...
with 3' recessed ends, the integration products result from both substrate cleavage and subsequent strand transfer. Consequently, integration could only occur after significant and correct 3' processing on viral ends. Here, it is therefore impossible to assess and quantitate the substrate specificity of strand transfer activity of HIV-1 IN which shows nearly no site-specific cleavage on heterologous substrates.

The integration patterns generated by MVV IN and HIV-1 IN with identical substrates are slightly different (compare Fig. 6a, lanes 11 and 12, with Fig. 6b, lanes 11 and 12). Moreover, incubation of MVV IN with CAEV U3 and MVV U3 also results in different product patterns (compare Fig. 6b, lanes 3 and 7). Therefore, retroviral IN proteins may have individual integration site preferences, as concluded previously (Craigie, 1992; Leavitt et al., 1992; Hong et al., 1993).

Disintegration activities of CAEV IN, MVV IN and HIV-1 IN

Besides endonucleolytic cleavage and strand transfer, IN proteins also catalyse a reversal of integration, called disintegration (Fig. 2c). This activity was tested using 43-mer dumbbell substrates (DBS) resembling intermediates of the ligation of a viral DNA (branch) into a target DNA (base) (Fig. 1b). All three integrases were found to exhibit optimal disintegration activities under the conditions already used for endonucleolytic cleavage and strand transfer. In disintegration reactions, IN proteins preferentially break the bond between the branch and the base of the substrate; meanwhile a new bond in the host part is generated in a one-step polynucleotidyl transfer reaction. Such assays result in radiolabelled 16-mers as depicted in Fig. 7(a) (after incubation in the presence of...
Mg$^{2+}$ and in Fig. 7(b) (after incubation with Mn$^{2+}$). The different mobilities of 16-mers resulting from CAEV and MVV U5 DBS and all other dumbbell substrates are probably due to differences in their base compositions (Frank & Köster, 1979). Comparing the results of disintegration and integration assays, it is noticeable that CAEV IN catalyses the disintegration with efficiencies similar to MVV IN (compare Fig. 7b, lanes 2 and 4 with lanes 6 and 8), while no integration activity was detectable testing various CAEV IN preparations on different substrates (Fig. 6a, b, lanes 2, 6, and 10).

In contrast to endonucleolytic cleavage and strand transfer, CAEV IN and MVV IN also show efficient disintegration if Mg$^{2+}$ is the divalent cation of choice (Fig. 7a, lanes 2, 4, 6 and 8). Under the same conditions, almost no products are found with HIV-1 IN (Fig. 7a, lanes 10 and 12). Again, more efficient activities are seen in the presence of Mn$^{2+}$ resulting not only in the expected 16-mer (DP$_{16}$) but in additional cleavage products (Fig. 7b). The most conspicuous of these fragments represents a 28-mer (DP$_{28}$, Fig. 7b) whose occurrence has not yet been documented for disintegration reactions with IN proteins from other sources. Its formation possibly results from a nucleophilic attack of the 3′-terminal OH group directed not against the phosphodiester bond between the conserved CA dinucleotide and the host part, but against the spatially most closely neighboured bond following 12 nucleotides downstream (compare Fig. 1b). The +1, −1 and −2 product bands seen above and below the 16-mers are thought to result from disintegration events concerning phosphodiester bonds up- and downstream of the conserved CA. A comparable ability of HIV-1 IN to catalyse transesterification at different sites of altered disintegration substrates was reported recently by Chow & Brown (1994).

**Substrate specificities of disintegration activities**

All IN proteins tested catalyse efficient disintegration not only on their cognate substrates (Fig. 7a, b), but also with each of the other U5 and U3 dumbbell oligonucleotides (data not shown). The expected 16-mer product is found after incubation of all enzyme–substrate combinations, whereas the additional 28-mer band occurs only after incubation with CAEV IN or MVV IN. Furthermore, the patterns of +1, −1 and −2 products generated on different substrates vary from enzyme to enzyme.

**Discussion**

Hexahistidine-tagged IN proteins of CAEV and MVV were expressed in E. coli and purified by Ni–NTA affinity chromatography. For direct comparisons, the integrase of HIV-1 was expressed in the same system and purified under identical conditions. To minimize possible effects of the tags and to avoid the co-purification of C-terminal degraded products, a C-terminal site of hexahistidine fusion was chosen for all three enzymes. It was thought that the highly conserved N-terminus would be much more sensitive to alterations than the variable C-terminal region of the enzyme, although it was reported previously that no differences in integrase activity occur with hexahistidine tails attached either to the N or C terminus (Drellich et al., 1992; Bushman & Wang, 1994). As such, no significant influences of the C-terminal fused histidine residues on CAEV IN and MVV IN activities were expected.

Comparing the expression levels of CAEV IN, MVV IN and HIV-1 IN, it was observed that the latter was synthesized in lower amounts. This is probably due to a less efficient usage of HIV-1 IN codons in E. coli (Holler et al., 1993) and to additional translation initiation from an internal ribosome-binding site (Hizi & Hughes, 1988). N-terminal sequencing proved that the truncated byproduct of HIV-1 IN expression is in fact translated from an internal AUG codon located downstream of a perfect Shine–Dalgarno sequence.

Like HIV-1 IN, CAEV IN and MVV IN were found to be active in both site-specific cleavage and disintegration, whereas only MVV IN and HIV-1 IN seemed to be able to integrate the cleaved substrate into an acceptor DNA. Specific endonucleolytic cleavage by CAEV IN and MVV IN resulted in the removal of two terminal nucleotides from the plus strand of the substrate, a finding that was described previously for IN proteins of other retroviruses (Katzman et al., 1989; Bushman et al., 1990; Craigie et al., 1990; Pahl & Flügel, 1993; Vink et al., 1994). In addition to the expected 18-mer product, a dominant n−1 band was detected in all reactions with HIV-1 IN. Since co-purification of an E. coli exonuclease due to protein–protein interactions cannot be ruled out with a native purification, different preparations and alternative purification procedures were tested to remove all possible contaminants from HIV-1 IN. Although these efforts failed, direct evidence that the formation of 19-mers is due to HIV-1 IN has yet to be found.

As described previously, HIV-1 IN cleaved its U5 substrate with a higher efficiency than its U3 substrate (Bushman & Craige, 1991; Engelman et al., 1991; Drellich et al., 1992; Sherman et al., 1992), whereas CAEV IN and MVV IN demonstrated comparable cleavage activities with both its U5 and U3 substrates. Recently, the preferential cleavage of U5 substrates was also reported for human foamy virus (HFV) (Pahl & Flügel, 1993) and feline immunodeficiency virus (FIV) (Vink et al., 1994). In HFV as well as in primate lentiviruses the 5′ ends of U3 each encode essential parts of an open reading frame, bel-3 and nef, respectively.
Therefore, competing functional necessities could result in a worse suitability of HFV and HIV-1 U3 sequences for endonucleolytic cleavage. On the other hand, U3 regions of MVV and CAEV overlap only a few C-terminal codons of env. Hence, the 5' terminal U3 sequences of animal lentiviruses are more likely to have evolved to function as good substrates for site-specific cleavage than the respective regions of HFV and HIV. Furthermore, in this examination the lower cleavage activity on HIV-1 U3 was clearly shown to be a property of the respective nucleotide sequence because CAEV IN and MVV IN also cleaved this substrate with reduced efficiencies. In spite of this plausible explanation for the different cleavage efficiencies on HFV and HIV-1 U5 and U3 substrates, no significant differences between these and corresponding substrates from other lentiviruses are observable. Only the CA dinucleotide and the following G residue are conserved among lentiviruses, whereas all other sequence positions are more or less variable with a preference for pyrimidine nucleotides at positions 1 and 6 to 9 (counting from the 3' end of the CA-containing strand). Moreover, FIV IN cleaves its U3 substrate less efficiently and less specifically than its U5 substrate, although its U3 sequences overlap only a few C-terminal codons of env. Further investigations are needed to explain these contradictory findings.

CAEV IN and MVV IN were found to have relaxed sequence requirements for site-specific cleavage. This was shown by efficient processing of all substrates investigated. In contrast, HIV-1 IN was barely active on CAEV and MVV substrates but it cleaved efficiently its cognate DNA ends. Again, a comparison of nucleotide sequences corresponding to the outer U5 and U3 termini of ovine–caprine lentiviruses and HIV-1 does not explain the reduced substrate specificity of CAEV IN and MVV IN. Although previous mutational analyses of HIV-1 IN substrates revealed that the conserved CA dinucleotide and some of its neighbouring residues are critical for specific cleavage (LaFemina et al., 1991; Leavitt et al., 1992; Sherman et al., 1992), the molecular basis of different processing activities is yet to be elucidated.

A remarkable discrepancy between the three IN proteins investigated in this study is the lack of detectable integration activity of CAEV IN. To exclude the possibility that this finding is due to the oligonucleotide sequences of its cognate substrates, CAEV IN was tested on MVV and HIV substrates, and conversely MVV IN and HIV-1 IN were tested on CAEV substrates. While HIV-1 IN showed integration activities only with its cognate substrates, MVV IN integrated all substrates including those of CAEV, although with different efficiencies. It is noticeable that MVV IN converts CAEV U3 more efficiently than its own U3 substrate, whereas no integration products occurred when CAEV IN was tested on substrates cognate for MVV and HIV-1. These observations clearly demonstrate that the CAEV substrates are suitable for strand transfer and that the lack of integration activity in the case of CAEV IN results only from properties of the enzyme.

Since it is unlikely that the IN protein of CAEV 75-G63 does not catalyse integration in vivo, the question arises whether limited domains of the bacterially expressed CAEV IN were artefactually damaged. For instance, the block might be due to a partial misfolding of the C-terminal domain resulting in an inability to bind non-cognate target DNA or to form multimers which are possibly required for DNA joining but not for the other activities. Both could completely abolish strand transfer without affecting site-specific cleavage or disintegration. Another explanation is offered by a possibly more stringent requirement for concerted two-end activity, as shown in the case of the murine leukaemia virus (MuLV) (Murphy & Goff, 1992). However, the results obtained with CAEV IN demonstrate that efficient disintegration as well as efficient site-specific cleavage can independently occur from DNA joining. Currently, alternative expression systems and purification schemes are being tested to overcome the apparent inability of CAEV IN to perform strand transfer.

Only slight differences between the three IN proteins were found regarding the respective disintegration activities. Although CAEV IN appeared unable to integrate any substrate investigated, the enzyme disintegrated the viral parts of all dumbbell substrates tested. Taking into account that disintegration cleavage occurs preferentially downstream of the CA dinucleotide, this activity might function with the same molecular mechanism as site-specific cleavage. Therefore, the release of the viral branch could result from a complete transesterification reaction as well as from an attack of the 3'-terminal OH group or exogenous nucleophiles such as water or glycerol. To clarify this, studies with altered substrates are already in progress.

In contrast to endonucleolytic cleavage and strand transfer, disintegration was found to be almost independent of the substrate origin. Only the patterns of by-products varied slightly between the different enzymes. These results are in agreement with previous findings that IN can promote disintegration even when the conserved CA is replaced by another dinucleotide or if the viral DNA part of the substrate is reduced to a single A nucleotide (Chow et al., 1992). The chromatographic differences between the 16-mer CAEV and MVV U5 DBS disintegration products and all other 16-mer disintegration products are suspected to be due to different base compositions (Frank & Köster, 1979).

Fundamental agreement exists comparing the properties of MVV IN from German strain 461 with data from
the IN protein of Icelandic strain 1514 (Katzman & Sudol, 1994). Both enzymes efficiently catalyse endonucleolytic cleavage, strand transfer and disintegration using cognate U5 and U3 substrates. Nevertheless, a striking difference between MVV 461 IN and MVV 1514 IN is the different requirement for divalent cations. As reported, the latter is absolutely dependent on Mn²⁺ to exhibit catalytic activities, whereas no enzyme functions were found using Mg²⁺. In contrast, MVV 461 IN catalyses site-specific cleavage in the presence of either of the two cations, albeit with significantly higher productivities in the presence of Mn²⁺. The same holds for the closely related IN protein of CAEV and the respective enzymes of avian leucosis virus (Katzman et al., 1989), RSV (Katz et al., 1990), HIV-1 (Drellich et al., 1992) and HIV-2 (Vink et al., 1991b). In addition, Mg²⁺ efficiently promotes the disintegration activities of both MVV 461 IN and CAEV IN, whereas there are almost no products with HIV-1 IN. Besides MVV 1514 IN, strong requirements for Mn²⁺ have also been reported for the integrases of MuLV (Craigie et al., 1990) and HFV (Pahl & Flügel, 1993). Whether these findings are due to the respective amino acid composition or possibly result from different expression systems, purification protocols or assay conditions remains to be clarified by further investigations.

In summary, the results discussed above demonstrate considerable differences as well as basic similarities between the IN proteins of ovine-caprine lentiviruses and HIV-1. Although these findings contribute to the knowledge of integrase functions, further investigations on CAEV IN and MVV IN are needed to answer the basic unsolved questions. In particular, elucidation of the three-dimensional structure of IN proteins would help to explain the molecular mechanisms of the observed integrase activities, but so far all efforts to obtain crystals of complete IN proteins for X-ray crystallographic analysis have failed. Therefore, it appears necessary to study and compare further IN proteins from different sources. A more detailed knowledge of naturally occurring integrase variants might help in the construction of an active chimeric IN protein with sufficient solubility for crystallization and consequent analysis.

We thank Otto Christian Straub for the kind provision of MVV 461 and FLT cells, Hans-Georg Kräusslich for pNL4-3, Susanne Kempter for protein sequencing, and Martina Gaßmann for helpful technical assistance. We also thank Achim Rzih, Armin Saalmüller and Lisa Szabo for critical reading of the manuscript.

This work was supported in part by a grant from the Bundesministerium für Forschung und Technologie (FG 5-1075).

References


Hizi, A. & Hughes, S. H. (1988). Expression of the Moloney murine...


