Nucleic Acids Research, 1993, Vol. 21, No. 6 1419–1425

Identification of the catalytic and DNA-binding region of the human immunodeficiency virus type I integrase protein

Cornelis Vink, Antoinette A.M.Oude Groeneger and Ronald H.A.Plasterk*
Division of Molecular Biology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

Received December 22, 1992; Accepted February 2, 1993

ABSTRACT

The integrase (IN) protein of the human immunodeficiency virus (HIV) is required for specific cleavage of the viral DNA termini, and subsequent integration of the viral DNA into target DNA. To identify the various domains of the IN protein we generated a series of IN deletion mutants as fusions to maltose-binding protein (MBP). The deletion mutants were tested for their ability to bind DNA, to mediate site-specific cleavage of the viral DNA ends, and to carry out integration and disintegration reactions. We found that the DNA-binding region resides between amino acids 200 and 270 of the 288-residues HIV-1 IN protein. The catalytic domain of the protein was mapped between amino acids 50 and 194. For the specific activities of IN, cleavage of the viral DNA and integration, both the DNA-binding domain and the conserved amino-terminal region of IN are required. These regions are dispensable however, for disintegration activity.

INTRODUCTION

Integration of double-stranded viral DNA is essential for replication of retroviruses. The integration reaction can be divided into two steps. In the first step, two nucleotides are removed from the 3′ termini of blunt-ended, linear viral DNA. In the second step, the retroviral DNA is inserted into the DNA of the infected cell. During this integration step, the recessed 3′ ends of the viral DNA are coupled to target DNA, resulting in a gapped intermediate. The bases at the 5′ ends of the viral DNA that are not base-paired are subsequently removed, and the single-stranded gaps repaired, presumably by cellular enzymes (for reviews, see references 1 and 2).

The only protein required for integration, is the viral integrase (IN) protein. integrase can site-specifically cleave the viral DNA ends directly 3′ of a conserved CA-3′ dinucleotide sequence (3–6). It was shown for HIV-1 and HIV-2 IN, that this cleavage reaction (or donor cut reaction) involves a nucleophilic attack of appropriate nucleophiles on the phosphodiesterbond 3′ of the CA-3′ sequence, resulting in the release of a dinucleotide (7, 8). IN subsequently mediates the coupling of the recessed 3′ ends of the viral DNA to target DNA (the DNA strand transfer reaction). This reaction most likely proceeds via a one-step mechanism that does not involve a covalent protein-DNA intermediate (7). HIV-1 IN has also been shown to mediate the reversal of the integration reaction, called disintegration (9).

The substrate requirements of IN have been determined (3, 10–13). it is still unclear, however, what the function is of the different parts of the protein. It has been found that IN contains two regions that are conserved among retroviral and retrotransposon IN proteins (14, 15). One of these regions, near the amino-terminus of IN, has the structure His-X_3-His-X_{20-30}-Cys-X_2-CYS (the first His being at position 12 in HIV-1 IN) and might be involved in coordination of a Zn^{2+} ion (14, 16). This region is referred to as the HHCC region. Point-mutational analyses of various IN proteins have shown that mutation of each of the conserved His and Cys residues in the HHCC region significantly reduce site-specific cleavage activity, and to a lesser extent integration activity (15, 17–20). Generally, disintegration activity was hardly affected by these mutations (17–19). On the basis of these results it was speculated that the HHCC region might be involved in the recognition and correct positioning of the viral DNA substrate (17–19). The HHCC region was, however, found to be dispensible for DNA-binding by IN (15, 21–23).

The most highly conserved region in retroviral and retrotransposon IN proteins is the middle part, extending from approximately amino acid 113 to 162 in the HIV IN proteins. This region, the so-called D(35)E region (15, 24), contains the invariable residues Asp116 and Glu152. Another invariable aspartic acid residue is located at position 64 of the IN proteins. Mutations in each of these 3 invariable residues resulted in loss of all activities of IN (17, 18, 25, 26), suggesting that they are part of the catalytic site that is involved both in site-specific DNA cleavage and in DNA strand transfer. The involvement of a single catalytic site in both cleavage and strand transfer is inferred from the common effect of most point mutations on both activities. Apart from its role in catalysis, the D(35)E region has also been suggested to be involved in DNA binding (15).

Since most of the studies on IN mutants concentrated on only a subset of IN activities, and because comparison between studies done for integrases from different retroviruses may be difficult,

* To whom correspondence should be addressed

we systematically investigated all known activities of HIV-1 IN in relation to the primary structure of the protein. An extensive series of amino- and carboxyl-terminal deletion mutants was generated, and tested for DNA-binding, site-specific DNA cleavage, DNA strand transfer and disintegration activity. We found that mutants that carry deletions of 50 amino acids at the amino-terminus and 94 amino acids at the carboxyl-terminus, are still able to mediate disintegration. These mutants however, were no longer capable of mediating site-specific cleavage and integration. We show here, that the loss of these activities by carboxyl-terminal IN deletion mutants is precisely correlated with a loss in DNA-binding activity. The DNA-binding domain of IN was found to reside within amino acids 200 to 270, which is a part of IN that does not coincide with the catalytic domain of the protein (amino acids 50 to 194). This shows that IN contains separable domains for DNA binding and catalysis.

MATERIALS AND METHODS

Construction of plasmids encoding IN deletion mutants

The HIV-1 IN gene was fused in-frame to the 3′ end of the *malE* gene in vector pMAL-c (New England Biolabs). The IN coding sequence was derived from plasmid pHXB-2D (27). The 5′ terminus of the IN coding sequence was changed by site-directed mutagenesis into the sequence shown in Figure 1. This alteration, which introduced a *Bam*HI endonuclease restriction site directly 5′ of the IN gene, does not change the amino acid sequence of the IN protein. In addition, a *Hin*dIII site was generated downstream of the *Sal*I site at position 5366 of the HIV-1 sequence (28). The IN gene was cloned into the *Bam*HI and *Hin*dIII sites of vector pMAL-c, generating plasmid pRP619.

Carboxyl-terminal deletion mutants of HIV-1 IN were generated by the introduction of 2 in-frame translation termination codons into the IN sequence via oligonucleotide-directed mutagenesis using the pMa/c system (29). The *Bam*HI/*Hin*dIII fragment from pRP619 that contains the IN gene was cloned into the *Bam*HI and *Hin*dIII sites of vector pMc5-19 (29), resulting in plasmid pRP290. Single-stranded DNA of pRP290 was isolated and annealed to the *Bam*HI/*Hin*dIII vector fragment of pMa5-19. The sequences of the oligonucleotides that were used for the mutagenesis were as follows:

*Spe*I
C-Δ9: 5′-CTGTCTACTTGCTTACTAGTCATCACCTGC-3′
*Spe*I
C-Δ18: 5′-GCCATCTGTTTTTACTAGTCCCTAATGATC-3′
*Spe*I
C-Δ32: 5′-TCTTGGCACTACTTACTAGTCACTATTATC-3′
*Nhe*I
C-Δ41: 5′-TCTTGTATTTATAGCTAGCCTTCACCTTTC-3′
*Nhe*I
C-Δ51: 5′-CCAGAGGAGTTATGGCTAGCCTTTCCAAAG-3′
*Spe*I
C-Δ66: 5′-CCTGTAATAAACTTACTAGTTTTGAATTTT-3′
*Spe*I
C-Δ81: 5′-TAATTCTTTAGTTTACTAGTCTGTTGCTAT-3′
*Spe*I
C-Δ94: 5′-ATTCTTTCCCCTTACTAGTACCCCCCAATC-3′
*Spe*I
C-Δ105: 5′-CCCTTTTCTTTTTTACTAGTGGATGAATAC-3′
*Spe*I
C-Δ117: 5′-GTACTGCTGTCTACTAGTGTTCAGCCTG-3′

Figure 1. The altered sequence at the 5′ end of the HIV-1 IN gene. Only the coding strand is shown 5′ to 3′ from left to right. The amino acids that are predicted to be encoded by the IN sequence are depicted by the single-letter code below the DNA sequence.

*Nhe*I
C-Δ125: 5′-AGCCTGATCTCATAGCTAGCCTATAATTTT-3′
*Spe*I
C-Δ133: 5′-TTTTCTTTAATTACTAGTTCATAGATTC-3′
*Nhe*I
C-Δ139: 5′-ATTCATAGATTATAGCTAGCCTTGACTTTG-3′

In each case 2 in-frame stop codons and a diagnostic endonuclease restriction site were introduced by these oligonucleotides. The mutated IN genes were cloned into vector pMAL-c using the *Bam*HI and *Hin*dIII restriction sites (Figure 1).

Several amino-terminal deletion mutants of IN were generated by deletion of specific restriction fragments from the 5′ end of the IN gene in plasmid pRP619. Mutant N-Δ5 was made by deletion of the *Bam*HI/*Cla*I fragment from pRP619 that is shown in Figure 1. The *Bam*HI and *Cla*I recessed ends were filled-in using Sequenase (United States Biochemical Corp.) and the resulting blunt ends were ligated. To generate mutant N-Δ23, pRP619 was digested with *Asp*718 and *Nhe*I (see Figure 1). The cohesive ends were subsequently filled-in and ligated. Mutant N-Δ33 was generated by digestion of pRP619 with *Asp*718 and *Nru*I, filling-in of the *Asp*718 protruding ends, and ligation. Mutant N-Δ38 was made by ligation of the 1.5 kilobasepairs (kb) *Pvu*II/*Hin*dIII fragment from pRP290 that contains the IN gene, to the *Hin*dIII and filled-in *Bam*HI termini of vector pMAL-c. The amino-terminal deletion mutants N-Δ50 and N-Δ72 were generated by oligonucleotide-directed mutagenesis using the following oligonucleotides:

*Bam*HI
N-Δ50: 5′-GTCCATGCATGGATCCTCCTTTTAGCTGAC-3′
*Bam*HI
N-Δ72: 5′-CTACCAGGATAACGGATCCTTCTAAATGTG-3′

The *Bam*HI site that was introduced by the oligonucleotides, and the *Hin*dIII site at the 3′ end of the IN gene, were used to clone the truncated genes into the *Bam*HI and *Hin*dIII sites of vector pMAL-c.

Fragments encoding the carboxyl-terminal region of HIV-1 IN were cloned as follows. The *Bfr*I fragment that encodes amino acids 172 to 279 was isolated from construct C-Δ9. The cohesive ends of this fragment were filled-in using Sequenase. The fragment was then cloned into vector pMAL-c, after digestion of the vector with *Asp*718, filling-in of the recessed ends and treatment with calf intestinal alkaline phosphatase (Boehringer Mannheim). The resulting construct (M-172-279) encodes amino acids 172 to 279 of HIV-1 IN fused to maltose-binding protein. To clone the fragment of the IN gene encoding amino acids

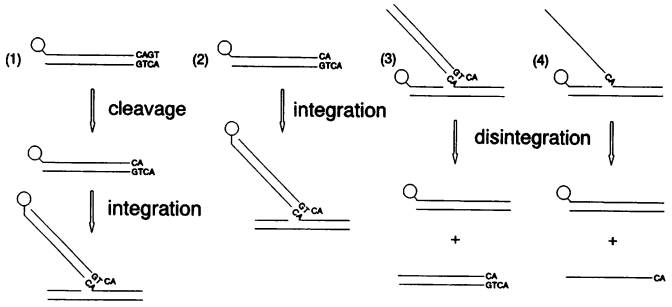

Figure 2. The oligonucleotide substrates that were used in this study, and the reactions mediated by IN. Substrate #1 represents the HIV-1 U5 end and has previously been described (13). The IN protein can specifically remove the two 3′ terminal nucleotides from one of the strands of this 28-bp substrate. The strand that is nicked by IN is indicated by an open circle (which denotes the presence of a ^{32}P label) at the 5′ end. The 5′ p-GT-OH 3′ dinucleotide that is removed during cleavage is not shown. Substrate #2 is used as substrate to determine integration activity. In the integration reaction, the 26-nucleotide strand is transferred to other DNAs, which in our assays are other substrate #1 or #2 molecules. Substrate #3 and #4 are used to test disintegration activity and have previously been described (18). The ^{32}P label in these substrates is present at the 5′ end of a 13-mer oligonucleotide strand. In disintegration, a trans-esterification takes place that couples the 3′-OH of the 13-mer to the phosphate-group directly 3′ of the conserved viral CA-3′ sequence. As a result, the viral DNA part of the substrate, which is either double-stranded (substrate #3) or single-stranded (substrate #4), is released, and a longer radiolabeled DNA strand of 29 nucleotides is generated.

200−279, a polymerase chain reaction (PCR) was performed using construct C-Δ9 as template and as primers oligonucleotide C-Δ9 (see above) and the following oligonucleotide:

*Bam*HI
AB3223: 5′-CACGGGGATCCATAGTAGACATAATAGCAAC-3′

The 0.2 kb PCR fragment was purified, digested with *Bam*HI and *Spe*I, and subsequently ligated into vector pMAL-c by using the *Bam*HI and *Xba*I restriction sites. The sequence of the cloned fragment was checked by dideoxy sequencing using the Sequenase system (United States Biochemical Corp.).

Expression and purification of HIV-1 IN fusion proteins

E.coli strain JM101 (30) was transformed with the various pMAL-c derivatives. Expression and purification of the fusion proteins was carried out essentially as described previously (26).

Oligonucleotide substrates

The synthetic oligonucleotides that were used in the cleavage, integration and disintegration assays are depicted in Figure 2. Labeling of the oligonucleotides with ^{32}P, subsequent purification and annealing have previously been described (13, 18).

Cleavage, integration and disintegration reactions

Cleavage, integration and disintegration mixtures (10 μl) contained 20 mM MOPS, pH 7.2, 50 mM NaCl, 3 mM $MnCl_2$, 3 mM DTT, 1 mM Tris, 0.05 mM EDTA, 0.1 mM β-mercaptoethanol, 4% glycerol, 0.2 pmoles of the appropriate oligonucleotide substrate (see Figure 2) and 20 pmoles of full-length fusion protein. Reactions were carried out at 30°C for 1 h, and were stopped by addition of 10 μl of formamide loading dye (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol). After incubation for 3 min at 80°C, 5 μl of the mixtures was loaded onto either a 12% (cleavage and integration reactions) or a 15% (disintegration reactions) denaturing polyacrylamide gel. After electrophoresis, reaction products were visualized by autoradiography. Integrase activity, except for site-specific cleavage activity (see the legend to Table 1), was quantified by densitometry, using an Ultrascan XL Enhanced laser densitometer (LKB).

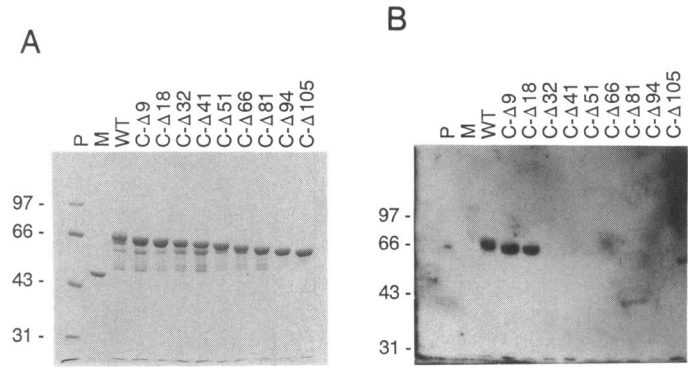

Figure 3. **(A)** SDS−PAGE analysis of several of the purified fusion proteins. The proteins were separated on a 7.5% denaturing polyacrylamide gel and detected by Coomassie-staining. **(B)** Southwestern blot analysis of the same proteins as shown in (A). Molecular weights of the marker proteins are depicted on the left in kilodaltons. P, molecular weight markers; M, MBP- β-galα; WT, wild-type HIV-1 IN fused to the carboxyl-terminus of MBP.

Southwestern blot analysis

Southwestern blot analysis was carried out as described by Woerner *et al.* (23). Proteins were separated on 7.5% denaturing polyacrylamide gels.

RESULTS

Purification of deletion mutant derivatives of HIV-1 IN

Wild-type HIV-1 IN and IN deletion mutants were expressed in *E.coli* as fusions to maltose-binding protein (MBP). All fusion proteins were purified in the same way by a one-step purification procedure (26). Several of the purified fusion proteins are shown by SDS−PAGE analysis in Figure 3A. The faster-migrating proteins that were copurified with the fusion proteins presumably represent degradation products of the full-length proteins, as has been noted previously (26). The carboxyl-terminal IN mutants with large deletions were less sensitive to protease digestion than the proteins with smaller deletions. The amino-terminal deletion mutants were approximately equally sensitive to digestion by proteases as wild-type fusion protein (data not shown). This suggests that the carboxyl-terminal region of IN is more accessible to proteases than the amino-terminal and central regions of IN. The relatively high sensitivity of the carboxyl-terminal region of non-fused HIV-1 IN to protease digestion has previously been observed (17). One of the mutants, C-Δ125, was not tested because of low expression of the protein.

The amino-terminal IN deletion mutants

The amino-terminal 50 amino acids of IN comprise one of the two regions that are conserved among retroviral IN proteins (14, 15). This region (the HHCC region) is a putative Zn^{2+}-binding

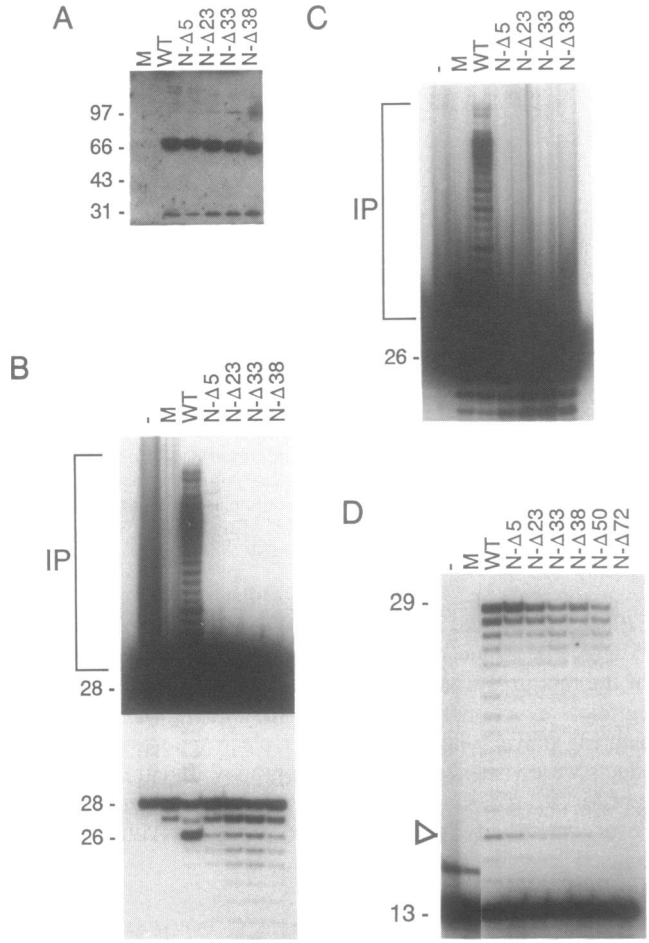

Figure 4. HIV-1 IN amino-terminal deletion mutants. **(A)** Southwestern blot analysis. M, MBP-β-galα. Molecular weight standards are shown on the left in kilodaltons. **(B)** Site-specific cleavage of substrate # 1 (Figure 2). The top panel is a longer exposure of the lower panel. **(C)** The integration activity of the deletion mutants was tested using substrate # 2 (Figure 2). IP, integration products **(D)** Disintegration. Substrate # 3 was used (Figure 2). The positions of the substrate (13 nucleotides) and the product (29 nucleotides) are indicated on the left. −, incubation in the absence of protein; M, MBP-β-galα. Lengths of oligonucleotides are depicted on the left in (B), (C) and (D). The triangle in (D) points to a band that is a presumed product of reintegration (see text).

Table I. The activities of deletion mutants of HIV-1 IN

Mutant	cl.	Int.	Dis.	DNA-b.
N-Δ5	−	+/−	++	++
N-Δ23	−	−[a]	+	++
N-Δ33	−	−[a]	+	++
N-Δ38	−	−[a]	+	++
N-Δ50	−	−[a]	+	++
N-Δ72	−	−	−	++
C-Δ9	++	++	++	++
C-Δ18	+	+	++	++
C-Δ32	−	−	+	−
C-Δ41	−	−	+	−
C-Δ51	−	−	+	−
C-Δ66	−	−	+/−	−
C-Δ81	−	−	+/−	−
C-Δ94	−	−	+/−	−
C-Δ105	−	−	−	−
C-Δ117	−	−	−	−
C-Δ125	ND	ND	ND	ND
C-Δ133	−	−	−	−
C-Δ139	−	−	−	−
M-171-279	ND	ND	ND	++
M-200-279	ND	ND	ND	++

The various activities were determined as described in Materials and Methods. cl., site-specific cleavage; Int., integration; Dis., disintegration; DNA-b., DNA-binding. ++, 50−100% of wild-type level; +, 10−50% of wild-type level; +/−, 1−10% of wild-type level; −, less than 1% of wild-type level; ND, not determined. The integration, disintegration and DNA-binding activities were quantified by densitometry; the level of site−specific cleavage activity was estimated from autoradiographs, and was scored negative when the band at the 26-mer position was not stronger than the 27-mer and not much stronger than the 25-mer. Mutant C-Δ125 was not tested because of low expression of the protein.

[a]Although no integrations were observed when we used a precleaved viral DNA, we found that these mutants are able to carry out reintegration of the viral DNA after it is released from the disintegration substrate (see text).

domain (14, 16). To investigate the function of the amino-terminus of IN in general, and the HHCC region in particular, we generated several amino-terminal deletion mutants of HIV-1 IN. One of these mutants, N-Δ5, probably still contains the HHCC region; the other mutants have a partially (N-Δ23, -Δ33, -Δ38) or totally deleted HHCC region (N-Δ50 and -Δ72). The deletion mutants were tested for all known activities of IN: DNA-binding, site-specific cleavage of viral DNA ends, integration of precleaved viral DNA ends, and disintegration.

We found that all amino-terminal mutants were able to bind DNA (Figure 4A and Table I): the HHCC region is apparently not required for DNA-binding in this assay.

Deletion of only 5 amino acids from the amino terminus of IN greatly decreased the activity of the protein both in site-specific cleavage and in integration assays (Figure 4 and Table I). That this mutant might possess (a low level of) specific cleavage activity could be inferred from the appearance of integration products after prolonged autoradiography of the gel (Figure 4B); removal of two nucleotides from the viral 3′ ends is required before integration can occur. Specific cleavage activity however, could not be detected due to nonspecific nuclease activity. Integration activity of mutant N-Δ5 could barely be observed when a precleaved oligonucleotide substrate was used (Figure 4C). This is due to lower sensitivity of the integration assay compared to the cleavage assay (compare the activity of wild-type IN in Figure 4B and 4C). The other amino-terminal mutants did not show any activity either in donor cut or integration reactions. The inactivity of mutants N-Δ23, -Δ33 and -Δ38 in donor cutting is not consistent with the conclusion of Drelich *et al.* (25), that an amino-terminal deletion mutant of HIV-1 IN that lacks amino acids 1 to 39 is still active in donor cutting. Rather, our findings are in agreement with the results of Engelman and Craigie (17) and of Vincent *et al.* (19), who found that the introduction of double point mutations in the HHCC region (H12>N/H16>N and H12>A/E13>L, respectively) totally abolishes site-specific cleavage activity.

Up to 50 amino acids can be deleted from the amino-terminus of IN without loss of the ability to carry out disintegration (Figure 4D and Table I). This shows that the HHCC region, which is deleted in mutant N-Δ50, is not required for catalysis per se, but is involved in reactions that, unlike the disintegration reaction (9), require specific recognition of the viral DNA ends, like the donor cut and integration reactions. The products that are visible in Figure 4D between the substrate and the full-length 29-base product of disintegration, probably represent products of nonspecific nuclease activity as well as reintegration products (see below). When 72 amino acids were deleted from the amino-

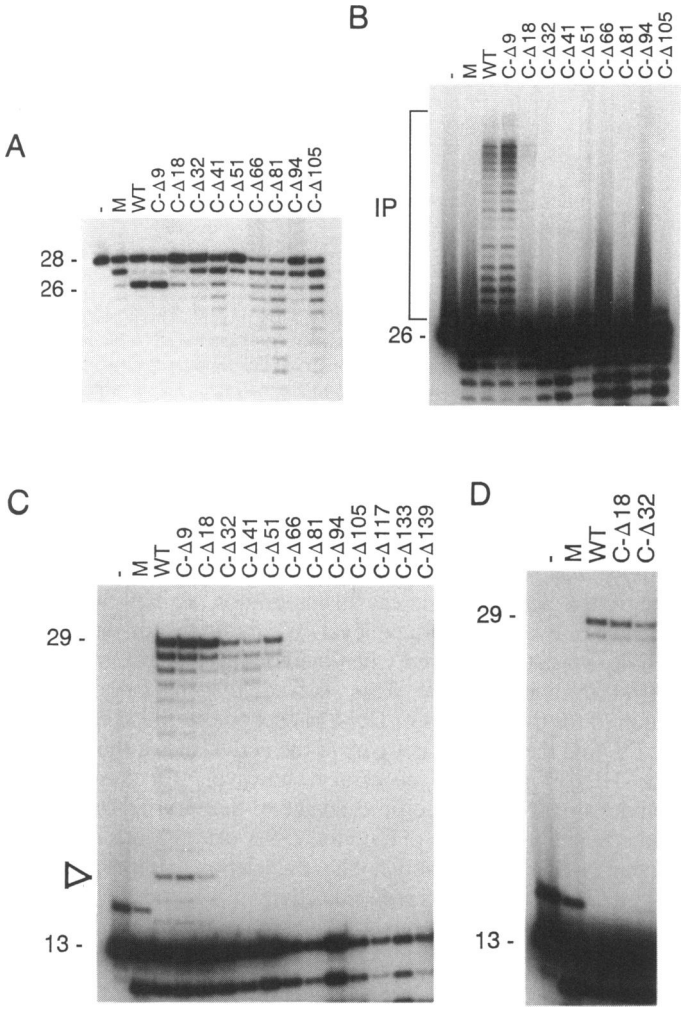

Figure 5. The carboxyl-terminal deletion mutants. **(A)** Site-specific cleavage using substrate #1 (Figure 2). **(B)** The integration reaction. Substrate #2 was used (Figure 2). IP, integration products. **(C)** Disintegration reactions with substrate #3 (Figure 2). The triangle points to a band that probably is a product of reintegration. **(D)** Disintegration reactions with substrate #4 (Figure 2). −, incubation without protein; M, MBP-β-galα. The lengths of the oligonucleotides are shown on the left.

terminus, disintegration activity could not be detected. This is consistent with the observation that the conserved aspartic acid residue at position 64 of HIV-1 IN, which is deleted in mutant N-Δ72, is required for the catalytic activities of IN, and might be part of the catalytic site of the protein (17, 18, 25, 26).

The carboxyl-terminal IN deletion mutants

The DNA-binding region of the integrase proteins of HIV-1 and Rous sarcoma virus has been shown to reside in the carboxyl-terminal half of the proteins (21, 23). To study the function of this region, which is not very conserved among retroviral and retrotransposon integrases, a series of carboxyl-terminal deletion mutants was made. As shown in Figure 3B up to 18 amino acids can be deleted from the carboxyl-terminus of IN without losing DNA-binding activity. DNA-binding can not be detected however, when 32 or more amino acids are deleted from the carboxyl-terminus. Therefore, the boundary of the DNA-binding region at the carboxyl-terminal side is presumably located between amino acids 256 and 270 of HIV-1 IN.

Analogous to what was found with the amino-terminal mutants, the ability of the carboxyl-terminal mutants to carry out site-specific cleavage corresponded with the ability to mediate integration. Mutant C-Δ18 was active in site-specific DNA cleavage as well as integration, albeit at a lower level than the wild-type fusion protein (Figure 5A and B and Table I). Both activities were lost when 32 or more amino acids were removed. For the carboxyl-terminal mutants a correlation was found between DNA-binding activity and the ability to carry out both donor cutting and integration. In contrast to this, disintegration activity was not absolutely restricted to deletion mutants that are able to bind DNA (Figure 5C). Disintegration activity could still be detected when as much as 94 amino acids were removed from the carboxyl-terminus. However, mutants that can still bind DNA had a higher disintegration activity than those that are unable to bind DNA. We conclude that the ability of IN to bind DNA does contribute to disintegration activity, but is not absolutely required.

Similarly as was found for the amino-terminal mutants, products are generated in the disintegration reactions with the carboxyl-terminal mutants that have a lower mobility than the substrate, but a higher mobility than the full-length 29-nucleotide disintegration product. It has been shown previously, that several of these products are the result of reintegration of the released viral DNA into larger radiolabeled DNA strands (19). An obvious candidate product of reintegration is the prominent band at the position of 16 nucleotides in Figure 4D and 5C. To test whether this product is indeed the result of a reintegration event, a disintegration substrate was generated of which the viral DNA part that is removed during disintegration is single-stranded (Figure 2). Vincent *et al.* have shown that this single-stranded viral DNA end can not be integrated by IN after disintegration (19). As shown in Figure 5D, the 16-nucleotide product was not observed when the alternative disintegration substrate was used, indicating that this product was indeed formed by reintegration of the viral DNA end. In Figure 5C the 16-mer reintegration product is not generated by mutants that do exhibit disintegration activity, but are not able to bind DNA. Interestingly, the amino-terminal mutants that lack all or part of the HHCC region are still able to carry out reintegration (compare Figure 4D and 5D). In summary, both the HHCC region and the DNA-binding region of IN are required for integration activity, but dispensible for disintegration activity. For reintegration activity, however, the DNA-binding region but not the HHCC region is necessary.

In all assays in Figure 5 a high level of nonspecific nuclease activity can be observed with the mutants that have carboxyl-terminal deletions larger than 41 amino acids. Whether this nonspecific activity is an intrinsic property of the deletion mutants, or whether a higher amount of contaminating *E.coli* nucleases are copurified with these mutants is not clear. However, all proteins behaved similarly during purification. Furthermore, a similar nuclease activity could be detected in purified preparations of a HIV-2 non-fusion deletion mutant that lacked 73 amino acids from the carboxyl-terminus (data not shown). Presumably, the large carboxyl-terminal deletions do not affect the ability of IN to cleave DNA, but do affect the specificity of this cleavage activity; as a result, the carboxyl-terminal deletions might have turned a site-specific endonuclease into a general endonuclease.

Taken together with the results from the amino-terminal deletion analysis, our findings show that the catalytic site of HIV-1 IN is located between amino acids 50 and 194. This region contains the invariable amino acids Asp64, Asp116 and Glu152, which have been shown to be essential for catalytic activity (17, 18, 25, 26).

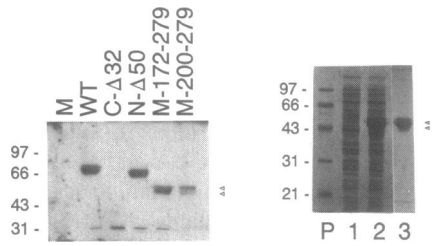

Figure 6. (A) Southwestern blot analysis of mutants M-172-279 and M-200-279. (B) SDS−PAGE analysis of a crude lysate of strain M-200-279 before (lane 1) and after induction with IPTG (lane 2). Lane 3, purified protein M-200-279. Proteins were separated on a 10% denaturing polyacrylamide gel, and detected by Coomassie-staining. The two bands that are present in the preparations are indicated with triangles. P, Molecular weight standards. Molecular weights are depicted on the left.

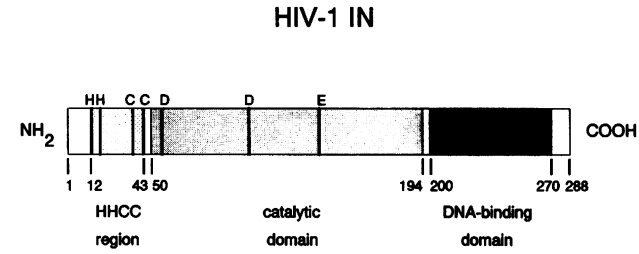

Figure 7. The different regions in the HIV-1 IN protein. The positions of the HHCC region, the catalytic region (residues 50−194) and the DNA-binding region (residue 200−279) are indicated. The conserved His residues at positions 12 and 16, and Cys residues at positions 40 and 43 are indicated. The invariable amino acids that are essential for IN activity (D64, D116 and E152), and are presumably part of the catalytic region are also shown.

The DNA-binding region of IN

To fine-map the amino-terminal border of the DNA-binding region of HIV-1 IN, we generated two fusion proteins: M-172-279, which contains amino acids 172 to 279 of HIV-1 IN fused to MBP, and M-200-279, containing residues 200−279 fused to the carboxyl-terminus of MBP. The ability of these proteins to bind DNA was tested by Southwestern analysis (Figure 6A). Both proteins exhibited DNA-binding, indicating that the DNA-binding domain is confined to the region from residue 200 to 270. In the preparation from mutant M-200-279 a double protein band can be seen (Figure 6A). These same bands were also observed by SDS−PAGE analysis and Coomassie-staining of both the crude lysate of the induced *E.coli* culture that expresses M-200-279 and the purified protein (Figure 6B). The sequence encoding HIV-1 IN amino acids 200 to 279 in strain M-200-79 was checked, but no abnormalities were observed (data not shown). Therefore, it is probable that these proteins are generated in *E.coli* by proteolysis of a single full-length fusion protein, giving rise to a full-length product that exhibits wild-type DNA-binding activity and a smaller product that only shows partial DNA-binding activity. In conclusion, our results indicate that the DNA-binding region (amino acids 200 to 270) and the catalytic region (amino acids 50 to 194) of IN are separable regions that can function independently.

DISCUSSION

To determine which parts of the HIV-1 IN protein are required for the various activities of the protein, we tested a series of amino- and carboxyl-terminal deletion mutants in several assays. The deletion mutants were expressed in *E.coli* as fusions to maltose-binding protein to enable a similar purification procedure for each deletion gene product. We found that a deletion of only five amino acids from the amino-terminus of IN already significantly inhibited both site-specific cleavage and DNA strand transfer activity. These activities were undetectable after deletion of the 23 amino-terminal residues. This deletion includes part of the conserved HHCC region which might be involved in the coordination of a Zn^{2+}-ion (14, 16). The importance of the HHCC region for IN activity has previously been demonstrated by point-mutational analyses of various IN proteins (15, 17−20). In general it was found, that mutations in either of the invariable His and Cys residues of the HHCC region gave rise to reduction of all known catalytic activities of IN. Site-specific cleavage activity was somewhat more affected by these mutations than integration activity, whereas disintegration activity was least affected or even of wild-type level. We found that disintegration activity was still observed when the complete HHCC region of HIV-1 IN was removed. This showed, that this region is not required for the catalysis of DNA phosphodiesterbond formation by IN, and therefore is not part of the active site of the protein. The HHCC region is necessary, however, for the specific activities of IN: site-specific cleavage of the viral DNA termini and efficient integration of the viral DNA (see below). Although mutant N-Δ5 does probably not have a deletion that includes (part of) the HHCC region, its reduced activity in site-specific cleavage as well as in integration might be due to a structural change that is caused by the deletion, which interferes with the proper folding of the HHCC region. It is also possible that the amino-terminal five residues are part of the HHCC region, or play a role that is independent of the function of the HHCC region. It has been suggested by Vincent *et al.* that the HHCC region is a separate DNA-binding domain that recognizes the viral DNA termini (19). In the light of our findings, however, it seems most likely that the HHCC region is not a DNA-binding domain by itself, but rather plays an assisting role in the recognition of the viral DNA ends. Deletion mutants that lack the carboxyl-terminal DNA-binding domain (residue 200 to 270; Figure 7) but contain both the catalytic domain (residue 50 to 194) and the HHCC region are neither able to detectably bind DNA nor to carry out a specific IN reaction.

For the carboxyl-terminal deletion mutants, the loss in both site-specific cleavage and integration activity correlated with the loss in DNA-binding activity. These activities were no longer detectable after deletion of 32 amino acids from the carboxyl-terminus. Disintegration activity, however, was still observed after removal of the 94 carboxyl-terminal residues of IN. Because the DNA-binding domain of the 288-amino acids HIV-1 IN protein was found to be located between residue 200 and 270, this implies that the ability of IN to bind DNA is not absolutely required in the disintegration reaction. This might be explained by a mechanism in which the protein hits the DNA substrate in a random fashion; once the substrate becomes properly positioned into the catalytic site of the protein, a DNA phosphodiester-linkage is made. It is also possible that several of the carboxyl-terminal deletion mutants (the ones that have a deletion that is smaller than 88 amino acids) contain a part of the DNA-binding

region that enables them to bind, albeit with reduced efficiency, to the disintegration substrate. This binding activity might not be detected by Southwestern analysis.

In contrast to the disintegration reaction, DNA-binding is required for reintegration: the integration of viral DNA ends that are released from the disintegration substrate by IN. Surprisingly, amino-terminal deletion mutants that have disintegration activity, but are not able to carry out integration of a precleaved viral DNA substrate at a detectable level (mutant N-Δ23 through to mutant N-Δ50), were all capable of mediating reintegration. A similar observation has previously been made for HIV-1 IN proteins with double and triple point mutations in the HHCC region (19). These results can be explained by the assumption that the HHCC region plays a role in the correct positioning of the viral DNA termini for specific cleavage and DNA strand transfer. A mutant that lacks the HHCC region might therefore not be able to properly position the viral DNA ends, which could be the rate-limiting step in site-specific cleavage and integration. The ability of this mutant to bind DNA however, might bring its catalytic site in the proximity of the reactive groups in the disintegration substrate, and promote a trans-esterification reaction. Because the viral DNA terminus that is released from the disintegration substrate might still be correctly positioned with respect to the catalytic site of the protein, it might be reintegrated by a second trans-esterification reaction.

In accordance with previous reports our findings are in favor of a model in which integrase contains a single active site (7, 8, 18, 31). The specific activities of IN (site-specific cleavage and integration) were similarly affected in all mutants. Our results also support, by the assignment of only one DNA-binding region to the IN protein, a model in which minimally two IN monomers are required per viral DNA end in the integration reaction; one monomer could bind the viral DNA end, whereas the other monomer could bind the target DNA.

ACKNOWLEDGEMENTS

We would like to thank Piet Borst, Sean Colloms, Chris Vos, Ramon Puras Lutzke and Dik van Gent for critical comments on the manuscript. We thank Dik van Gent for construction of plasmid pRP619. We thank Cindy Gerhardt at the Free University of Amsterdam for assistance with the densitometric analyses. This research was supported by grant 900-502-119 and grant PGS 82-189 from the Dutch Organization for Scientific Research. C.V. is currently supported by a research grant from Glaxo Group Research Limited (UK).

REFERENCES

1. Brown,P.O. (1990) *Curr. Top. Microbiol. Immunol.* **157**, 19–48.
2. Grandgenett,D.P. and Mumm,S.R. (1990) *Cell* **60**, 3–4.
3. Bushman,F.D. and Craigie,R. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 1339–1343.
4. Craigie,R., Fujiwara,T. and Bushman,F. (1990) *Cell* **62**, 829–837.
5. Katz,R.A., Merkel,G., Kulkosky,J., Leis,J. and Skalka,A.M. (1990) *Cell* **63**, 87–95.
6. Sherman,P.A. and Fyfe,J.A. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 5119–5123.
7. Engelman,A., Mizuuchi,K. and Craigie,R. (1991) *Cell* **67**, 1211–1221.
8. Vink,C., Yeheskiely,E., van der Marel,G.A., van Boom,J.H. and Plasterk,R.H.A. (1991) *Nucleic Acids Res.* **19**, 6691–6698.
9. Chow,S.A., Vincent,K.A., Ellison,V. and Brown,P.O. (1992) *Science* **255**, 723–726.
10. LaFemina,R.L., Callahan,P.L. and Cordingley,M.G. (1991) *J. Virol.* **65**, 5624–5630.
11. Leavitt,A.D., Rose,R.B. and Varmus,H.E. (1992) *J. Virol.* **66**, 2359–2368.
12. Sherman,P.A., Dickson,M.L. and Fyfe,J.A. (1992) *J. Virol.* **66**, 3593–3601.
13. Vink,C., van Gent,D.C., Elgersma,Y. and Plasterk,R.H.A. (1991) *J. Virol.* **65**, 4636–4644.
14. Johnson,M.S., McClure,M.A., Feng,D.-F., Gray,J. and Doolittle,R.F. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 7648–7652.
15. Khan,E., Mack,J.P.G., Katz,R.A., Kulkosky,J. and Skalka,A.M. (1990) *Nucleic Acids Res* **19**, 851–860.
16. Burke,C.J., Sanyal,G., Brunert,M.W., Ryan,J.A., LaFemina,R.L., Robbins,H.L., Zeft,A.S., Russell Middaugh,C. and Cordingley,M.G. (1992) *J. Biol. Chem.* **267**, 9639–9644.
17. Engelman,A. and Craigie,R. (1992) *J. Virol.* **66**, 6361–6369.
18. van Gent,D.C., Oude Groeneger,A.A.M. and Plasterk,R.H.A. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 9598–9602.
19. Vincent,K.A., Ellison,V., Chow,S.A. and Brown,P.O. (1993)??? *J. Virol.* in press.
20. LaFemina,R.L., Schneider,C.L., Robbins,H.L., Callahan,P.L., LeGrow,K., Roth,E., Schleif,W.A. and Emini,E.A. (1992) *J. Virol.* **66**, 7414–7419.
21. Mumm,S.R. and Grandgenett,D.P. (1991) *J. Virol.* **65**, 1160–1167.
22. Schauer,S. and Billich,A. (1992) *Biochem. Biophys. Res. Commun.* **185**, 874–880.
23. Woerner,A.M., Klutch,M, Levin,J.G. and Marcus-Sekura,C.J. (1992) *AIDS Res. and Hum. Retroviruses* **8**, 297–304.
24. Fayet,O., Ramond,P., Polard,P., Prere,M.F. and Chandler,M. (1990) *Mol. Microbiol.* **4**, 1771–1777.
25. Drelich,M.R., Wilhelm,R. and Mous,J. (1992) *Virology*, **188**, 459–468.
26. Kulkosky,J., Jones,K.S., Katz,R.A., Mack,J.P.G. and Skalka,A.M. (1992) *Mol. Cell. Biol.* **12**, 2331–2338.
27. Fisher,A.G., Collalti,E., Ratner,L., Gallo,R.C. and Wong-Staal,F. (1985) *Nature* **316**, 262–265.
28. Ratner,L., Haseltine,W., Patarca,R., Livak,K.J., Starcich,B., Josephs,S.F., Doran,E.R., Rafalski,J.A., Whitehorn,E.A., Baumeister,K., Ivanoff,L., Petteway,S.R.Jr, Pearson,M.L., Lautenberger,J.A., Papas,T.S., Ghrayeb,J., Chang,N.T., Gallo,R.C. and Wong-Staal,F. (1985) *Nature* **313**, 277–284.
29. Stanssens,P., Opsomer,C., McKeown,Y., Kramer,W., Zabeau,M. and Fritz,H.J. (1989) *Nucleic Acids Res.* **17**, 4441–4454.
30. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
31. Mizuuchi,K. (1992) *J. Biol. Chem.* **267**, 21273–21276.