Nucleic Acids Research, 1993, Vol. 21, No. 15 3373–3377

Identification of amino acids in HIV-2 integrase involved in site-specific hydrolysis and alcoholysis of viral DNA termini

Dik C.van Gent, 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 May 11, 1993; Revised and Accepted June 17, 1993

ABSTRACT

The human immunodeficiency virus integrase (HIV IN) protein cleaves two nucleotides off the 3′ end of viral DNA and subsequently integrates the viral DNA into target DNA. IN exposes a specific phosphodiester bond near the viral DNA end to nucleophilic attack by water or other nucleophiles, such as glycerol or the 3′ hydroxyl group of the viral DNA molecule itself. Wild-type IN has a preference for water as the nucleophile; we here describe a class of IN mutants that preferentially use the 3′ hydroxyl group of viral DNA as nucleophile. The amino acids that are altered in this class of mutants map near the putative active-site residues Asp-116 and Glu-152. These results support a model in which multiple amino acid side-chains are involved in presentation of the (soluble) nucleophile. IN is probably active as an oligomeric complex, in which the subunits have non-equivalent roles; we here report that nucleophile selection is determined by the subunit that supplies the active site.

INTRODUCTION

After reverse transcription of the HIV genomic RNA, the viral cDNA is integrated into the human chromosome. Integration is catalyzed by the viral IN protein. IN first removes two nucleotides from the blunt-ended viral DNA (1, 2), and subsequently couples the newly formed 3′ hydroxyl ends to phosphate groups in the target DNA (3–5).

Site-specific cleavage of the viral DNA is probably a one-step reaction: the phosphodiester bond 3′ of the conserved CA dinucleotide near the viral DNA end is exposed to nucleophilic attack (figure 1; ref. 6, 7). The natural nucleophile is most likely water, resulting in the release of a dinucleotide with a 5′ phosphate and a 3′ hydroxyl group. *In vitro* however, other nucleophiles can also be used, such as the alcohol glycerol, resulting in formation of a linear dinucleotide with a glycerol group coupled to the 5′ phosphate group (this reaction is called alcoholysis; ref. 7), or the 3′ hydroxyl group of the viral DNA, resulting in formation of a circular dinucleotide in which the 3′ hydroxyl group is coupled to the 5′ phosphate group (6).

Mutational analysis identified three candidate active-site residues in the HIV IN protein: Asp-64, Asp-116, and Glu-152 (8–12). These residues are conserved in all retroviral IN proteins (13, 14); proteins in which any one of these residues is mutated are inactive. Mutation of residues around Asp-116 and Glu-152 resulted in reduced activity. The active site is most likely contained in a polypeptide of amino acids 50 to 186 (15, 16); deletion mutants comprising these residues are able to mediate disintegration (the reverse of integration; ref. 17). One mutation in this region (Asn-117→Gln of HIV-1 IN) was found to cause a preference for the 3′ hydroxyl end of viral DNA as the nucleophile in the donor cut reaction (9). We here report the characterization of a whole class of mutants of HIV-2 IN (mapping around Asp-116 and Glu-152) that have the characteristic that they use the 3′ viral DNA end instead of water (or glycerol) as nucleophile. When substrates are presented that do not allow the formation of a circular dinucleotide, or when reaction conditions are chosen that do not support the formation of circular dinucleotides (Mg^{2+} instead of Mn^{2+}), the cleavage activity of these mutants is severely reduced.

The C-terminus of IN contains a DNA-binding domain, located between amino acids 200 and 270 (16), which is required for site-specific cleavage and integration, but not disintegration. IN mutants that are inactivated by mutations in different domains can complement each other (19,20), suggesting that IN is active as an oligomer (21) in which the subunits play non-equivalent roles. For example, the C-terminal DNA-binding domain and the central catalytic domain can be supplied on different molecules, suggesting that one subunit binds (viral) DNA, and another mediates cleavage of that DNA. We found that nucleophile selection is determined by the subunit that contains the intact active site.

MATERIALS AND METHODS

DNA techniques

DNA techniques were as described (22). Site-directed mutagenesis of the HIV-2 IN gene was done as described (12). Double mutant CΔ73/Q148L was made by site-directed mutagenesis of the construct expressing CΔ73, using oligonucleotide

* To whom correspondence should be addressed

5'-ACTACTCCTAGGCTCTGTG-3'. Mutant D64V/Q148L was made by cloning the *Msc*I-*Hind*III fragment carrying the mutation Q148L into the *Msc*I and *Hind*III sites of the construct expressing D64V.

Oligonucleotide cleavage reactions were carried out as described with substrates that were either labeled at the 3' end (7), or the 5' end (12). The reaction products were separated by denaturing polyacrylamide gelelectrophoresis and visualized by autoradiography. Autoradiographs were scanned using an Ultrascan XL Enhanced Laser Densitometer (LKB).

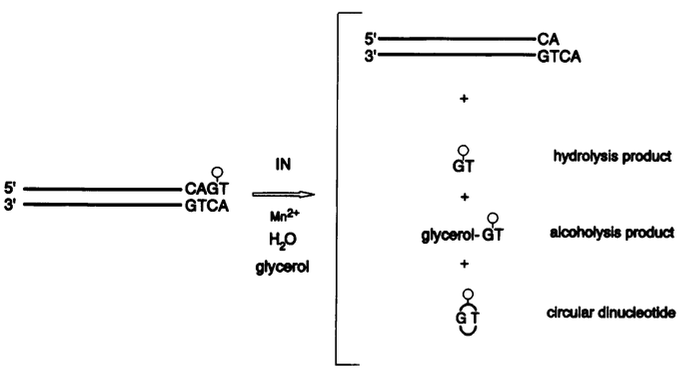

Figure 1. Schematic representation of site-specific cleavage by HIV IN. The position of the ^{32}P label is depicted by a circle.

Protein expression and purification

HIV-2 IN protein was expressed in *E.coli* and purified as described (12). Complementation assays were carried out as described (21).

RESULTS

Mutants with a different preference for nucleophiles

HIV IN cleaves two nucleotides off the 3' ends of viral DNA by exposing the phosphodiester bond to nucleophilic attack by water, glycerol, or the 3' hydroxyl group of the viral DNA (resulting in the formation of a linear dinucleotide, a linear dinucleotide with a glycerol group covalently coupled to the 5' phosphate group, or a circular dinucleotide, respectively; figure 2A lane 3).

We searched for mutants that have a different preference for nucleophiles. We therefore constructed and assayed a set of 31 HIV-2 IN mutants in which one amino acid had been changed (most of them described in ref. 12). We found that most mutants produced similar relative amounts of the three cleavage products (figure 2B). However, several mutants produced relatively more circular dinucleotides (figure 2A and table 1). We will refer to these mutants as alcoholysis mutants. The most striking example is mutant Q148L, which produces almost exclusively circular

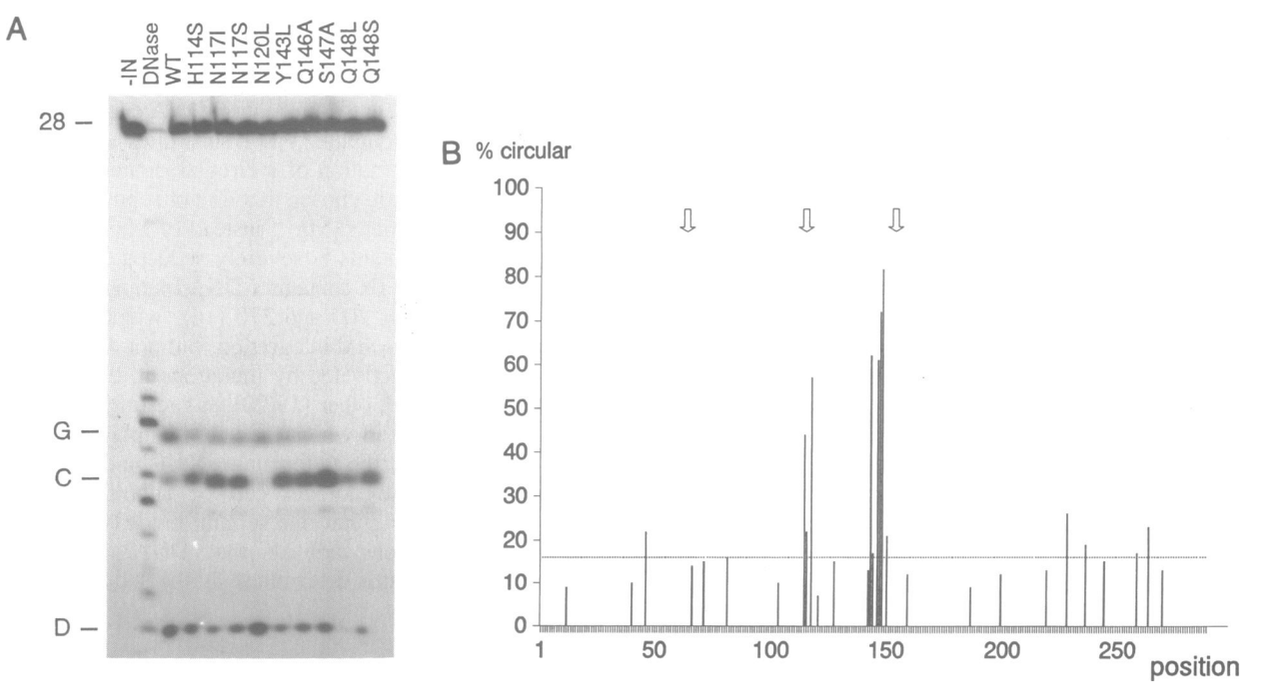

Figure 2. Alcoholysis mutants. (**A**) Cleavage reactions using oligonucleotide substrates labeled at the 3' end. The positions of the 28-mer substrate (28), glycerol adduct (G), circular dinucleotide (C), and (linear) dinucleotides (D) are indicated on the left. The various HIV-2 IN mutants used are indicated above the lanes. −IN = without IN, DNase = DNaseI degradation ladder of the oligonucleotide used in the cleavage reactions, WT = wild-type HIV-2 IN. The nomenclature of mutants is as follows: the first letter signifies the residue in wild-type HIV-2 IN, followed by its position and the amino acid to which it has been changed. (**B**) Diagram showing the relative amounts of circular dinucleotides produced by the various HIV-2 IN mutants. On the X-axis the position of the mutation is indicated, and on the Y-axis the amount of circular dinucleotides as percentage of the total amount of cleavage products. The stippled line signifies the relative amount of circular dinucleotides produced by wild-type HIV-2 IN. The positions of the putative active-site residues Asp-64, Asp-116 and Glu-152 are depicted by arrows. The following mutants were tested: H12L, C40S, K46L, T66A, K71L, S81A, K103Q, H114S, T115V, N117I, N117S, N120L, K127T, P142V, Y143L, N144V, Q146A, S147A, Q148L, Q148S, V150S, K159V, K186Q, R199S, K219Q, R228I, K236S, K244S, K258L, R263S, R269I. When two substitutions at the same position were tested, the average amount of circular dinucleotides is shown in the diagram.

dinucleotides (91% of the total amount of cleavage products, compared to 16% for wild-type HIV-2 IN. The cleavage activity of this mutant (*i.e.* the total amount of linear dinucleotides, glycerol adducts, and circular dinucleotides produced) is lower than that of wild-type HIV-2 IN. However, most other mutants of this class are not significantly less active than wild-type protein, but produce a higher amount of circular dinucleotides. These mutations therefore do not affect the ability for donor cleavage *per se*, but only the choice of the nucleophile. We found mutations in six different residues that resulted in production of a relatively higher amount of circular dinucleotides. Two of these amino acids map around the active-site residue Asp-116 (His-114 and Asn-117), and four are clustered close to the active-site residue Glu-152 (Tyr-143, Gln-146, Ser-147, and Gln-148).

The majority of the amino acid substitutions originally investigated were from hydrophilic to hydrophobic residues. We originally changed Asn-117 and Gln-148 to the hydrophobic amino acids isoleucine and leucine respectively. To investigate whether hydrophobicity of this region determines the choice of nucleophiles, we also changed Asn-117 and Gln-148 to the polar amino acid serine. We found that substitution by hydrophobic residues resulted in production of a somewhat higher percentage of circular dinucleotides than substitution by the polar amino acid serine, but in both cases there was a marked increase in the relative amount of circular dinucleotides compared to wild-type. This shows that hydrophobicity of these residues is not the main factor determining the choice of nucleophile.

Wild-type HIV-2 IN produced approximately twice as much linear as glycerol product (57% versus 27%; table 1). Some mutations resulted in a somewhat different ratio between these two products (*e.g.* mutant S147A produces 5 times more linear dinucleotides than glycerol products; table 1).

Alcoholysis mutants have a reduced site-specific cleavage activity when formation of circular dinucleotides is impossible

All experiments described above have been done under conditions that allow use of the viral 3′ hydroxyl group as nucleophile. We investigated whether mutants that have a preference for the viral 3′ hydroxyl group as nucleophile would use soluble nucleophiles under conditions that do not allow use of the 3′ hydroxyl end of viral DNA. This can be accomplished by changing the assay conditions (substituting Mg^{2+} for Mn^{2+}), or use of substrates that do not allow formation of circular dinucleotides (*e.g.* substrates in which one or three nucleotides are cleaved off; see also ref. 7).

When Mn^{2+} is replaced by Mg^{2+}, wild-type HIV-2 IN cleaves the viral DNA end, but does not form circular dinucleotides and formation of the glycerol adduct is less efficient (compare figure 3A lanes 3 and 4). Similarly, the alcoholysis mutants do not form circular dinucleotides in a buffer containing Mg^{2+}, and they produce somewhat less of the other products than in Mn^{2+} (shown for mutants N117S and Q148S in figure 3A lanes 6 and 8), resulting in a much reduced overall cleavage level.

Another way to prevent formation of circular dinucleotides is changing the number of nucleotides that are cleaved off. We therefore also tested substrates with either one, two, or three nucleotides 3′ of the conserved CA dinucleotide. Wild-type HIV-2 IN efficiently cleaves all three substrates (figure 3B lane 2). However, the 3′ hydroxyl end of substrates with one or three nucleotides 3′ of the CA dinucleotide can not be used as a nucleophile (7). Again, the cleavage level of the alcoholysis mutants was reduced, although less severely than in the cleavage reactions with Mg^{2+} (figure 3B).

In conclusion, we identified a class of mutants that preferentially use the 3′ hydroxyl group of the viral DNA as nucleo-

Table 1. Point mutants in the catalytic domain.

mutant	total	linear	glycerol	circular
wild type	100	57	27	16
H114S	79	39	16	44
N117I	75	21	18	60
N117S	79	31	15	54
N120L	92	75	18	7
Y143L	80	21	17	62
Q146A	87	28	11	61
S147A	126	23	5	72
Q148L	29	6	3	91
Q148S	54	21	5	74

The autoradiograph of Figure 1A was scanned and the amounts of linear dinucleotides, glycerol products and circular dinucleotides were determined. The second column shows the total amount of cleavage products as percentage of cleavage products produced by wild-type HIV-2 IN, columns 3, 4, and 5 show the amount of linear dinucleotides, glycerol products, and circular dinucleotides, expressed as percentage of the total amount of cleavage products.

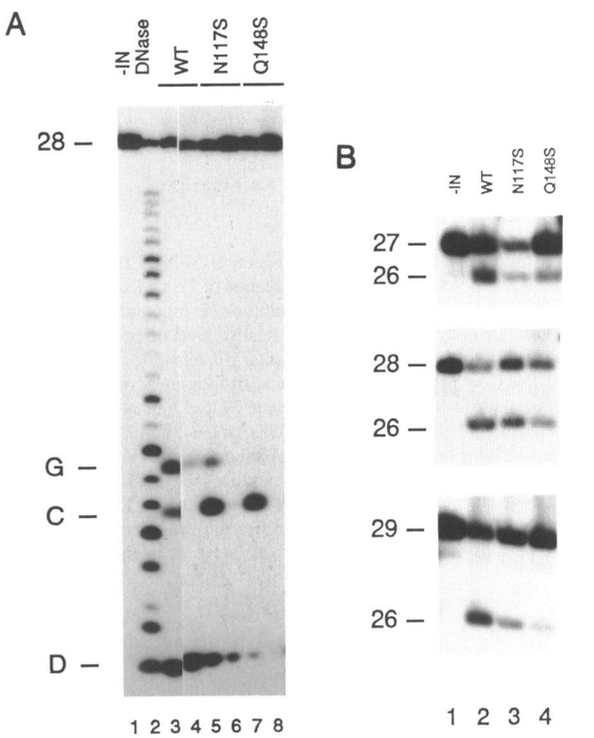

Figure 3. Cleavage under conditions that do not allow formation of circular dinucleotides. (**A**) Cleavage in the presence of Mn^{2+} (lanes 3, 5 and 7) or Mg^{2+} (lanes 4, 6 and 8) using wild-type HIV-2 IN (lanes 3 and 4), and the mutants N117S (lanes 5 and 6) and Q148S (lanes 7 and 8). Lane 1 shows a reaction without IN, and lane 2 contains a DNaseI degradation ladder of the oligonucleotide. The nomenclature is as in figure 1. (**B**) Cleavage of oligonucleotide substrates that contain one (upper panel), two (middle panel) or three nucleotides (lower panel) 3′ of the conserved CA-dinucleotide. Cleavage reactions without IN (lane 1), with wild-type HIV-2 IN (lane 2), and the mutants N117S (lane 3) and Q148S (lane 4) are shown. Substrates were labeled at the 5′ end, the positions of the substrates (29, 28, or 27) and the 26-mer cleavage products are indicated.

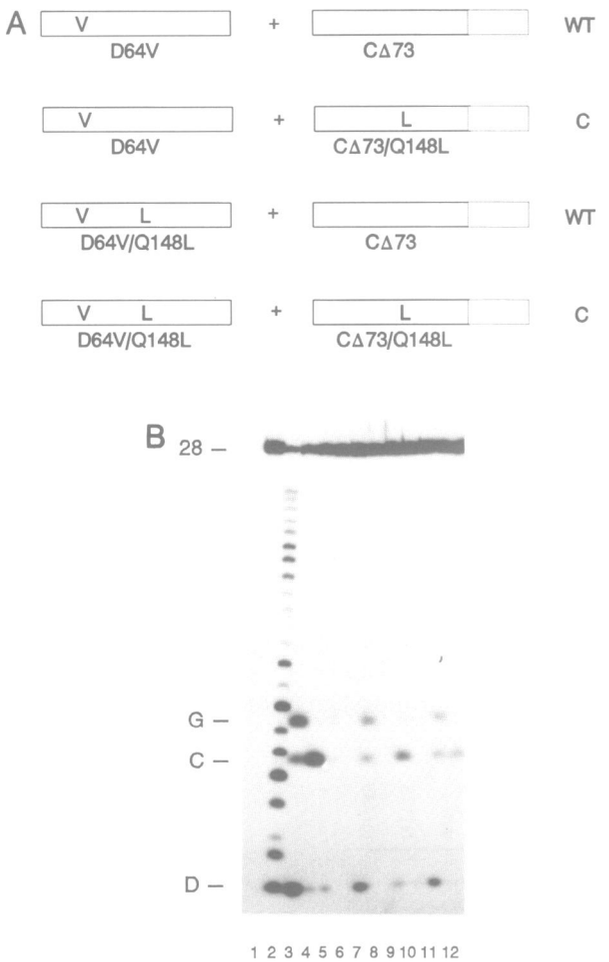

Figure 4. Complementation reactions between HIV-2 IN mutants containing alcoholysis mutations. (**A**) Schematic representation of the various complementation reactions. V = mutation Asp-64→Val (active-site mutation), L = mutation Gln-148→Leu (alcoholysis mutation). The relative level of circular dinucleotides is indicated; WT = relative amount of circular dinucleotides is similar to wild-type HIV-2 IN, C = primarily circular dinucleotides are formed. (**B**) Cleavage reactions containing wild-type HIV-2 IN (lane 3), the mutants Q148L (lane 4), D64V (lane 5), CΔ73 (lane 6), CΔ73/Q148L (lane 8), and D64V/Q148L (lane 10), and the combinations of D64V and CΔ73 (lane 7), D64V and CΔ73/Q148L (lane 9), D64V/Q148L and CΔ73 (lane 11), and D64V/Q148L and CΔ73/Q148L (lane 12). Lane 1 contains a reaction without IN, and lane 2 a DNaseI degradation ladder of the oligonucleotide. A slash signifies the presence of two mutations in the same protein molecule; further nomenclature as in figure 1.

phile in the cleavage reaction, and have a reduced cleavage activity when this nucleophile can not be used.

The nucleophile is selected by the subunit that provides the active site

HIV IN proteins mutated in different domains can complement each other (19, 20). Presumably, mutant CΔ73 (in which the C-terminal 73 amino acids have been deleted) contains the active site, but does not bind DNA, whereas mutant D64V binds DNA, but does not contain an intact active site. Both mutants alone do not mediate site-specific cleavage. A mixture of these mutants, however, is active: linear dinucleotides, glycerol adducts, and circular dinucleotides are produced in similar relative amounts as in the cleavage reaction with wild-type HIV-2 IN (figure 4 lane 7).

Mutant Q148L produces more than 90% circular dinucleotides. When the Gln-148→Leu mutation is present in both the C-terminal deletion mutant and the active-site mutant, the mixture of these mutants produces almost exclusively circular dinucleotides (figure 4 lane 12). To investigate whether the choice of nucleophile is determined by the subunit that provides the active site or by the one that contains the DNA-binding domain, we mixed the C-terminal deletion mutant and the active-site mutant in which the alcoholysis mutation was present only in the deletion mutant (figure 4 lane 9) or the active-site mutant (lane 11). When the alcoholysis mutation was present on the subunit that provides the active site (in this case the C-terminal deletion mutant), the mixture produced primarily circular dinucleotides, whereas the other combination produced the three cleavage products in similar amounts as wild-type HIV-2 IN, showing that the nucleophile is selected by the subunit that provides the active site.

DISCUSSION

IN exposes a specific phosphodiester bond near the viral DNA end to nucleophilic attack by water, glycerol, or the viral 3′ hydroxyl end. We here report the characterization of a class of mutant IN proteins that produce more circular dinucleotides (the result of nucleophilic attack by the 3′ hydroxyl group) and less of the other products. In addition, their cleavage activity in the presence of Mg^{2+} is much reduced, presumably because substitution of Mn^{2+} by Mg^{2+} prevents the formation of circular dinucleotides.

A large variety of nucleophiles can be used in the IN-catalyzed cleavage reaction, including glycerol, 1,2-ethanediol, L-serine, water and the 3′ hydroxyl group of the viral DNA molecule (6,7). This variety of nucleophiles might suggest that IN only exposes the phosphodiester bond and that it does not recognize the nucleophile. However, a common feature of these nucleophiles is that they have a hydroxyl group, which is used to attack the phosphodiester bond. We here show that substitution of several different amino acid residues leads to a change in the preference for nucleophiles. The main difference is observed between the soluble nucleophiles (water and glycerol) and the 3′ hydroxyl group of the viral DNA molecule. It is conceivable that soluble nucleophiles must be presented by IN, whereas the 3′ hydroxyl group of the bound viral DNA is present near the phosphodiester bond anyway. This would imply that the alcoholysis mutants specify amino acids involved in presentation of the nucleophile. Even mutants in which the size or the chemical properties of the side-chain of the substituted amino acid are similar to those of the wild-type residue (such as mutants S147A in HIV-2 IN or Asn-117→Gln in HIV-1 IN (9)), can lead to dramatic changes in the choice of nucleophiles. This suggests that the changed preference for nucleophiles is not caused by large changes in the three-dimensional structure. These residues might be involved in direct interactions with the nucleophile (thereby positioning and/or activating it), as has been proposed for the exonuclease action of *E.coli* DNA-polymerase I (23). Alternatively, they might form a cleft into which the nucleophile must enter before it can reach the phosphodiester bond. Small changes in the composition of such a cleft might alter nucleophile selection. We favor the first hypothesis, because nucleophiles that vary significantly in size and chemical properties can attack the specific phosphodiester bond.

In most mutational analyses cleavage assays were done in the presence of unphysiologically high concentrations of Mn^{2+}

(instead of Mg^{2+}). Substitution of Mg^{2+} for Mn^{2+} results in a high preference for water over glycerol, and under these conditions the 3′ hydroxyl group is not used at all. As expected, we found that the alcoholysis mutants then exhibit a very low level of donor cut, suggesting that their preference for the 3′ hydroxyl group is caused by difficulties in recruiting soluble nucleophiles. In other words, these mutants identify amino acids that are probably required for efficient cleavage *in vivo*, where primarily Mg^{2+} is present as divalent cation. The integration efficiencies and target-site preferences of most mutants described here are not altered (data not shown), suggesting that these residues are only involved in presentation of soluble nucleophiles.

Several oligomeric proteins have a shared active site. For example, the Flp recombinase in yeast provides three key catalytic residues (that probably activate a specific phosphodiester bond) on one subunit, and the active-site tyrosine residue (that leads the nucleophilic attack) on separate subunit (24). We show that the HIV IN subunit that provides the catalytic domain also determines the choice of the nucleophile, which is in agreement with the hypothesis that IN contains a single active site that is not shared between different subunits.

ACKNOWLEDGEMENTS

We thank Kees Vink and Sean Colloms for critically reading the manuscript. This work was supported by grant 88033 from the Dutch Ministry of Health (RGO) and a PIONIER grant from the Netherlands Organization for Scientific Research (NWO).

REFERENCES

1. Katzman,M., Katz,R.A., Skalka,A.M., and Leis,J. (1989) J. Virol., **63**, 5319–5327.
2. Sherman,P.A., and Fyfe,J.A. (1990) Proc. Natl. Acad. Sci. USA, **87**, 5119–5123.
3. Bushman,F.D., and Craigie,R. (1991) Proc. Natl. Acad. Sci. USA, **88**, 1339–1343.
4. Craigie,R., Fujiwara,T., and Bushman,F.D. (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. Engelman,A., Mizuuchi,K., and Craigie,R. (1991) Cell, **67**, 1211–1221.
7. 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.
8. Drelich,M., Wilhelm,R., and Mous,J. (1992) Virology, **188**, 459–468.
9. Engelman,A., and Craigie,R. (1992) J. Virol., **66**, 6361–6369.
10. Kulkosky,J., Jones,K.S., Katz,R.A., Mack,J.P.G., and Skalka,A.M. (1992) Mol. Cell. Biol., **12**, 2331–2338.
11. Leavitt,A.D., Shiue,L., and Varmus,H.E. (1993) J. Biol. Chem., **268**, 2113–2119.
12. van Gent,D.C., Oude Groeneger,A.A.M., and Plasterk,R.H.A. (1992) Proc. Natl. Acad. Sci. USA, **89**, 9598–9602.
13. Fayet,O., Ramond,P., Polard,P., Prère,M.F., and Chandler,M. (1990) Mol. Microbiol., **4**, 1771–1777.
14. Khan,E., Mack,J.P.G., Katz,R.A., Kulkosky,J., and Skalka,A.M. (1991) Nucleic Acids Res., **19**, 851–860.
15. Bushman,F.D., Engelman,A., Palmer,I., Wingfield,P., and Craigie,R. (1993) Proc. Natl. Acad. Sci. USA, **90**, in press.
16. Vink,C., Oude Groeneger,A.A.M., and Plasterk,R.H.A. (1993) Nucleic Acids Res., **21**, 1419–1425.
17. Chow,S.A., Vincent,K.A., Ellison,V., and Brown,P.O. (1992) Science, **255**, 723–726.
18. Schauer,M., and Billich,A. (1992) Biochem. Biophys. Res. Comm., **186**, 874–888.
19. Engelman,A., Bushman,F.D., and Craigie,R. (1993) EMBO J., in press.
20. van Gent,D.C., Vink,C., Oude Groeneger,A.A.M., and Plasterk,R.H.A. (1993) EMBO J., in press.
21. Jones,K.S., Coleman,J., Merkel,G.W., Laue,T.M., and Skalka,A.M. (1992) J. Biol. Chem., **287**, 16037–16040.
22. Sambrook,J., Fritch,E.F., and Maniatis,T. (1989) Molecular cloning: A laboratory manual, second edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
23. Beese,L.S., and Steitz,T.A. (1991) EMBO J., **10**, 25–33.
24. Chen,J.W., Lee,J. and Jayaram,M. (1992) Cell, **69**, 647–658.