The EMBO Journal vol.12 no.8 pp.3269-3275, 1993

Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex

Alan Engelman, Frederic D.Bushman[1] and Robert Craigie

Laboratory of Molecular Biology, National Institute of Diabetes, Digestive and Kidney Diseases, Bethesda, MD 20892 USA
[1]Present address: Infectious Disease Laboratory, Salk Institute, La Jolla, CA 92037, USA

Communicated by D.J.Sherratt

HIV-1 integrase protein possesses the 3′ processing and DNA strand transfer activities that are required to integrate HIV DNA into a host chromosome. The N-, C-terminal and core domains of integrase are necessary for both activities *in vitro*. We find that certain pairs of mutant integrase proteins, which are inactive when each protein is assayed alone, can support near wild type levels of activity when both proteins are present together in the reaction mixture. This complementation implies that HIV-1 integrase functions as a multimer and has enabled us to probe the organization of the functional domains within active mixed multimers. We have identified a minimal set of functional integrase domains that are sufficient for 3′ processing and DNA strand transfer and find that some domains are contributed in *trans* by separate monomers within the functional complex.

Key words: acquired immune deficiency syndrome/DNA recombination/human immunodeficiency virus/integration/multimeric complex

Introduction

Efficient retroviral replication requires the integration of a DNA copy of the viral RNA genome into a chromosome of the infected cell. Following reverse transcription, the 3′ ends of the linear viral DNA are cleaved to expose the CA_{OH} ends that become joined to chromosomal DNA. These processed 3′ ends are inserted into host DNA by a pair of DNA strand transfer reactions; the 5′ ends of the viral DNA remain unjoined in the resulting integration intermediate. Repair of the intermediate, which is presumably mediated by host enzymes, generates the final integrated product. For recent reviews on retroviral DNA integration see Goff (1992), and Whitcomb and Hughes (1992).

Purified integrase proteins of several retroviruses have been shown to carry out both the 3′ processing and DNA strand transfer reactions required for integration. *In vitro* reactions utilize DNA substrates that mimic the viral DNA ends. Human immunodeficiency virus type 1 (HIV-1) integrase both cleaves substrate DNA 3′ of the conserved CA (Sherman and Fyfe, 1990; Bushman and Craigie, 1991; Vink *et al.*, 1991a) and inserts the processed ends into a second DNA substrate molecule that serves as a target DNA in the strand transfer reaction (Bushman *et al.*, 1990; Bushman and Craigie, 1991; Lafemina *et al.*, 1991). HIV-1 integrase also promotes an apparent reversal of the strand transfer reaction, a process termed disintegration (Chow *et al.*, 1992).

The chemical mechanisms of the 3′ processing and DNA strand transfer reactions have been investigated *in vitro*. The stereochemical course of the reactions catalyzed by HIV-1 integrase has been followed by incorporating phosphorothioate of known chirality in substrate DNAs and determining the chirality of the phosphorothioate in reaction products. The chirality of phosphorothioate participating in either 3′ processing or strand transfer was found to invert during the course of these reactions. These results support the notion that each reaction takes place by a one-step transesterification mechanism and are consistent with a model in which the polynucleotidyl transfer reactions mediated by integrase are catalyzed by a common active site (Engelman *et al.*, 1991).

Determinants of HIV integrase important for function have been identified by studying the activities of purified mutant derivatives *in vitro*. Deletion mutagenesis has shown that both the N- and C-termini of HIV-1 integrase are important for 3′ processing and DNA strand transfer (Drelich *et al.*, 1992), although the biochemical roles of these regions in these reactions are unknown. HIV integrase has also been extensively studied by determining the activities of purified mutant proteins containing substitutions of conserved amino acid residues. The protein contains two amino acid sequence motifs that are conserved among retroviral and retrotransposon integrases (Khan *et al.*, 1991; Engelman and Craigie, 1992; Kulkosky *et al.*, 1992). One motif consists of two His and two Cys residues and is located near the N-terminus of the protein (Johnson *et al.*, 1986). Mutant proteins containing substitutions of these residues are, in general, more defective for 3′ processing and strand transfer than for disintegration activity (Engelman and Craigie, 1992; van Gent *et al.*, 1992; Leavitt *et al.*, 1993; Vincent *et al.*, 1993). The second motif is also conserved in the transposases of certain bacterial transposons; it consists of three acidic amino acid residues in the arrangement 'D,D-35-E' and is located within the relatively protease-resistant central core region of HIV-1 integrase. Mutant proteins containing certain substitutions of the conserved D,D-35-E residues are defective for all *in vitro* activities, suggesting that these residues are critical for catalysis of polynucleotidyl transfer (Drelich *et al.*, 1992; Engelman and Craigie, 1992; Kulkosky *et al.*, 1992; Lafemina *et al.*, 1992; van Gent *et al.*, 1992; Leavitt *et al.*, 1993). In addition, the core alone can promote the disintegration reaction, indicating that this domain probably contains the active site for all polynucleotidyl transfer activities (Bushman *et al.*, 1993).

Here we demonstrate that certain combinations of mutant HIV-1 integrase proteins, which individually exhibit little or no 3′ processing and strand transfer activities, efficiently

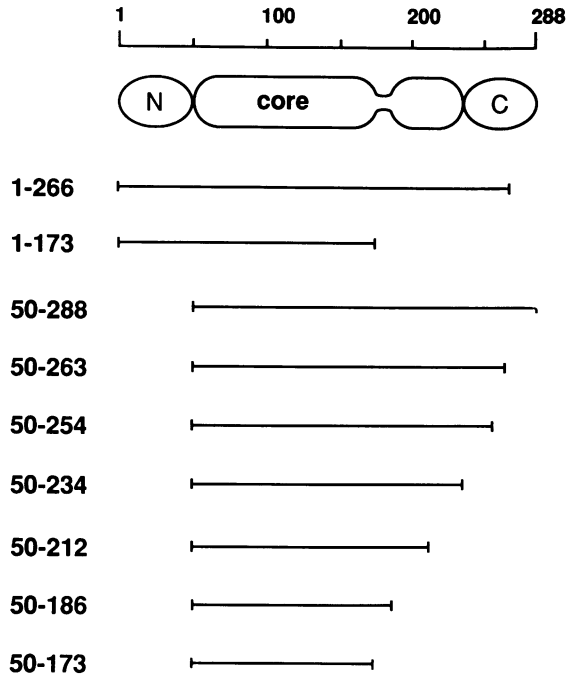

Fig. 1. Functional domains of HIV-1 integrase. The 288 amino acid residue wild type protein is shown depicting the domains required for 3′ processing and DNA strand transfer which are described in the text. We show that the N-terminal domain (residues 1–49) functions in *trans* to the core domain (residues 50–186). The region between residues 187 and 234 is important in *cis* to the core domain. The C-terminal region (residues 235–288) can function either in *cis* or *trans* to the core. Deletion mutants used in this study are indicated. The wild type residues retained in each mutant are noted on the left, alongside a line depicting the size of each protein relative to the wild type protein. N, N-terminal domain; C, C-terminal region; core, central core domain.

catalyze both reactions. We conclude that for both 3′ processing and DNA strand transfer, HIV-1 integrase functions as a multimer in which each integrase monomer requires only a subset of functional domains. In particular, an active core domain need not be present on the polypeptide that supplies both the N- and C-terminal functions required for 3′ processing and DNA strand transfer.

Results

HIV-1 integrase can function as a multimer in assays for 3′ processing and DNA strand transfer

HIV-1 integrase contains a core domain that is relatively resistant to proteolysis (Engelman and Craigie, 1992). Analysis of the biochemical activities of deletion derivatives of integrase has shown that this domain alone can promote the disintegration reaction, demonstrating that it is sufficient for catalysis of polynucleotidyl transfer (Bushman *et al.*, 1993). However, the 3′ processing and DNA strand transfer reactions require in addition both the N- and C-terminal parts of the protein (Drelich *et al.*, 1992; Bushman *et al.*, 1993). The roles of these regions have not yet been determined. In order to probe their roles, and the multimeric organization of integrase that is active in the 3′ processing and strand transfer reactions, we analyzed the ability of pairs of mutant proteins that each lack one or more of the above regions (Figure 1) to promote these reactions.

HIV-1 integrase mutants lacking the N-terminal region (IN^{50-288}) or part of the C-terminal region (IN^{1-266})

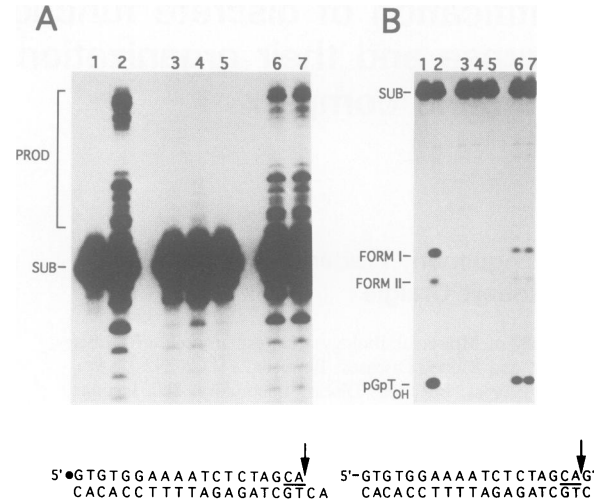

Fig. 2. Functional complementation of mutant HIV-1 integrase proteins *in vitro*. (**A**) Denaturing polyacrylamide gel of DNA strand transfer reactions. Integrase protein was omitted from the reaction in lane 1; lane 2 contained wild type integrase; lane 3, IN^{50-288}; lane 4, IN^{1-266}; lane 5, IN^{1-173}; lane 6, mixture of IN^{50-288} and IN^{1-266}; lane 7, mixture of IN^{50-288} and IN^{1-173}. The migration positions of the DNA substrate and strand transfer reaction products are marked SUB and PROD, respectively. The substrate, which models the terminal 21 bp of the viral U5 end, is shown in the lower half of the panel. The substrate is 'preprocessed' and therefore assays for strand transfer activity independent of 3′ processing activity (Bushman and Craigie, 1991). The arrow points to the 3′ end that is joined to target DNA in the strand transfer reaction products; a second substrate molecule serves as target in this reaction. A dot marks the position of ^{32}P label. (**B**) Denaturing polyacrylamide gel of 3′ processing reactions. The migration position of the substrate DNA is marked SUB. The migration positions of three dinucleotide cleavage products are marked form I, form II, and $pGpT_{OH}$. The form I product results from nucleophilic attack by glycerol, the form II product results from nucleophilic attack by the 3′-OH of substrate DNA, and the $pGpT_{OH}$ product results from hydrolysis (Engelman *et al.*, 1991; Vink *et al.*, 1991b). The reactions contained the same protein reactants as in panel A. The DNA substrate, which assays for the cleaved products of the 3′ processing reaction, is shown in the lower half of the panel. The position of cleavage is marked by the arrow. The DNA is labeled with ^{32}P 5′ of the T base that is highlighted in bold type. The conserved CA dinucleotide is underlined in both panels A and B.

displayed ~2 and 5% of wild type activity, respectively, in assays for DNA strand transfer (Figure 2A, lanes 3 and 4). Similar levels of activity were observed for these mutants in assays for 3′ processing (Figure 2B, lanes 3 and 4). However, when these proteins were both present in the reaction mixture, ~50% of wild type integrase strand transfer and 3′ processing activities were observed (Figure 2A and B, lane 6). Similar results were obtained from reactions containing a mixture of IN^{1-173} and IN^{50-288} (Figure 2A and B, lane 7). We conclude that these mixtures of mutant proteins are functional for 3′ processing and DNA strand transfer due to the formation of active mixed multimers and that the N- and C-termini of integrase need be present on only some subunits of the multimers. Figure 3 presents a summary of the activities of selected mixtures of mutant HIV-1 integrase proteins.

Physical evidence for HIV-1 integrase multimerization

The results of the previous experiments are consistent with the notion that HIV-1 integrase functions as a multimer in assays for 3′ processing and DNA strand transfer.

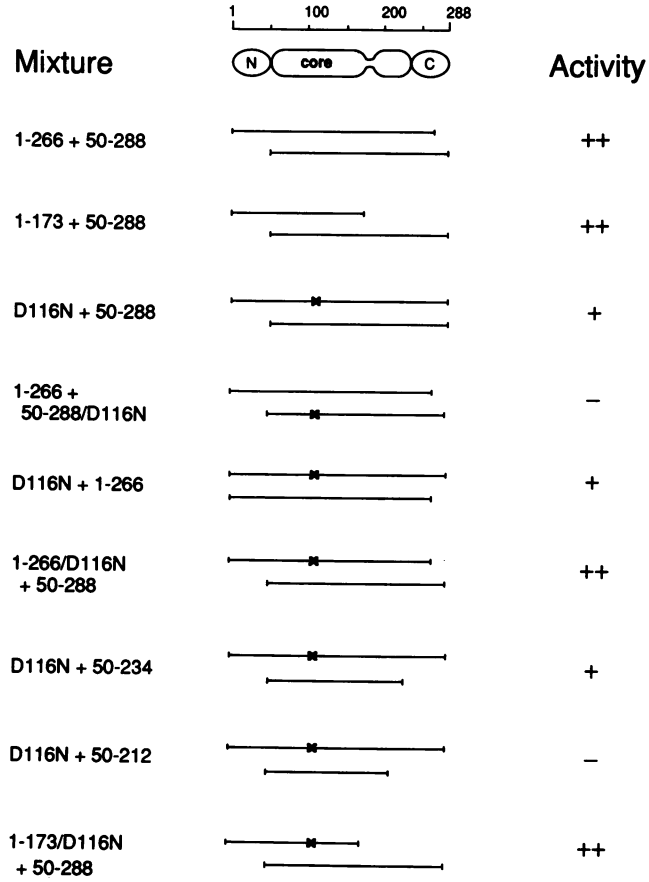

Fig. 3. Summary of selected mixtures of mutant HIV-1 integrase proteins assayed for 3′ processing and DNA strand transfer activities. The pairs of lines show the N- and C-terminal borders of the mutant proteins relative to wild type integrase (top). DNA strand transfer activity is given as the percent of wild type activity for the average of two experiments: ++, 20–50%; +, 10–20%; –, ≤5%. Mutants 1–266 and 50–288 individually displayed activities of 5 and 2%, respectively. All other mutants individually displayed activities of ≤0.5%. The relative levels of activities in 3′ processing assays were essentially the same as the values listed here for DNA strand transfer, both when the proteins were assayed individually and as mixtures. X marks the location of the D116N amino acid substitution.

Protein–protein cross-linking was subsequently used as a physical means to study the multimerization of HIV-1 integrase.

Incubation of wild type integrase with the cross-linker disuccinimidyl suberate (DSS), followed by SDS–PAGE, revealed a variety of cross-linked species (Figure 4A). The predominant product of intermolecular cross-linking migrated with the expected molecular weight of an integrase dimer. A less prominent band displayed the gel mobility expected for an integrase tetramer (Figure 4A).

We also analyzed the products of cross-linking deletion derivatives of integrase. Incubation of either IN^{1-266} or IN^{50-288} with DSS resulted in dimer formation, but cross-linked species with the mobilities expected of tetramers of either protein were not detected (data not shown). Similarly, dimers were the major product of cross-linking IN^{50-212} (Figure 4B, lanes 1 and 2). IN^{50-173} also formed cross-linked dimers, along with a variety of species of higher molecular weight (Figure 4B, lanes 3 and 4). These results suggest that sequences capable of dimerization lie within the core domain (residues 50–186) of HIV-1 integrase.

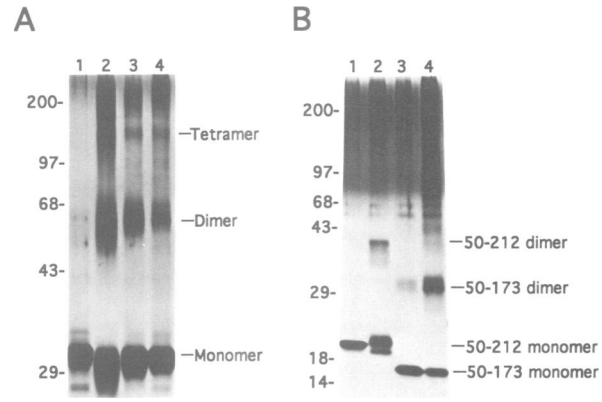

Fig. 4. Wild type integrase and deletion derivatives containing the core domain form dimers by protein–protein cross-linking. **(A)** SDS–polyacrylamide gel of wild type integrase and cross-linked products. DSS was omitted from the reaction in lane 1; lane 2 contained 200 μg/ml DSS; lane 3, 40 μg/ml DSS; lane 4, 8 μg/ml DSS. Products were electrophoresed in a 10% polyacrylamide gel. The migration position of wild type integrase is marked Monomer. The migration positions of cross-linked species, the sizes predicted for dimers and tetramers of integrase are marked. The migration positions of molecular weight standards in kDa are noted. **(B)** SDS–polyacrylamide gel of IN^{50-212}, IN^{50-173} and cross-linked products. Reactions in lanes 1 and 2 contained IN^{50-212}. Reactions in lanes 3 and 4 contained IN^{50-173}. DSS was omitted from the reactions in lanes 1 and 3. The migration positions of protein monomers and cross-linked dimers are marked. Products were electrophoresed in a 4–20% SDS–polyacrylamide gradient gel. A fraction of the IN^{50-173} protein was not completely reduced prior to electrophoresis; this material migrated at the dimer position in lane 3. The migration positions of molecular weight standards in kDa are noted.

The N-terminus of HIV-1 integrase acts in trans to the core domain that catalyzes polynucleotidyl transfer

The results presented in Figure 2 show that only one mutant integrase in a mixed multimer, for example of IN^{1-266} and IN^{50-288}, need have an intact N-terminal region to efficiently complement for 3′ processing and DNA strand transfer activity. Does the N-terminal region of IN^{1-266} function in *cis* with the core domain of IN^{1-266}, or in *trans* with the core of IN^{50-288}? To address this question similar experiments were carried out, except that one of the proteins was substituted by an integrase mutant containing a point mutation within the core domain. The choice of point mutation was based on the *in vitro* activities of two previously characterized mutant integrases, D116N and E152Q. Both of these proteins are defective for 3′ processing, strand transfer and disintegration activities due to a single amino acid substitution of one of the conserved acidic residues of the D,D-35-E sequence motif (Engelman and Craigie, 1992).

Reactions containing either D116N or E152Q together with IN^{50-288} resulted in an ~10-fold increase over the activity of either protein alone (~20% of the wild type activity) in both 3′ processing and strand transfer assays (Figure 5A and B, lanes 5 and 6). This demonstrates that in a functional multimer only one of two mutant proteins requires a core domain capable of polynucleotidyl transfer and that the N-terminal region (residues 1–49) of integrase can act in *trans* to that core. In order to test whether the N-terminus can also function in *cis* to the active core, reactions were carried out with IN^{1-266} and the double mutant $IN^{50-288/D116N}$. Only the basal activity of IN^{1-266} was observed (Figure 5A and B, lane 7). The lack of

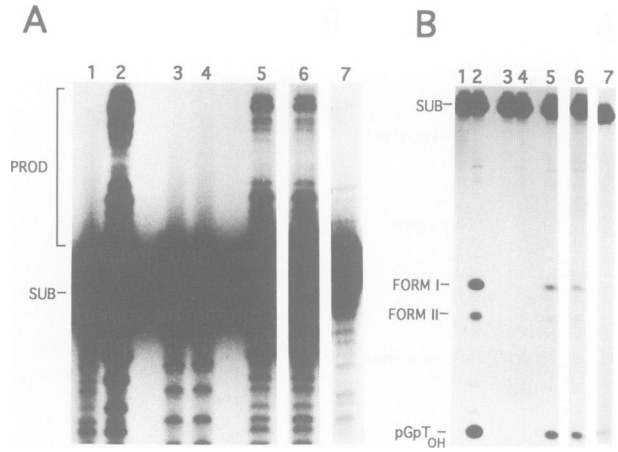

Fig. 5. The N-terminus functions in *trans* to the active core in mixed multimers. (**A**) Denaturing polyacrylamide gel of DNA strand transfer products. Integrase was omitted from the reaction in lane 1; lane 2 contained wild type integrase; lane 3, D116N; lane 4, E152Q; lane 5, mixture of D116N and IN^{50-288}; lane 6, mixture of E152Q and IN^{50-288}; lane 7, mixture of IN^{1-266} and $IN^{50-288/D116N}$. Other labeling is the same as in Figure 2A. The reaction in lane 7 was run in a separate polyacrylamide gel, along with a reaction containing wild type integrase. The relative activity of this mutant mixture is presented in Figure 3. (**B**) Denaturing polyacrylamide gel of 3′ processing products. The reactions contained the same protein reactants as in panel A. The reaction in lane 7 was from a separate polyacrylamide gel. Other labeling is the same as in Figure 2B.

complementation between these two mutants was not due to the absence of the C-terminus from IN^{1-266}, since a mixture of IN^{1-266} and D116N displayed similar activitites as the mixture of IN^{50-288} and D116N (Figure 3; see below). These results demonstrate that the N-terminal and core regions of IN^{1-266} are unable to function in *cis*. We conclude that for efficient 3′ processing and DNA strand transfer, an intact N-terminus and a core that is capable of polynucleotidyl transfer are required on different constituent monomers of the mixed multimer.

The C-terminus of HIV-1 integrase can function either in cis or trans to the active core

Having shown that the N-terminal region of integrase functions only in *trans* to the active core in a mixed multimer, we next probed the role of the C-terminal region. A mixture of either D116N or E152Q with IN^{1-266} resulted in an increase in the 3′ processing and DNA strand transfer activities compared with reactions containing either protein alone (Figure 6A and B, lanes 3 and 4; Figure 3). This result indicates that the C-terminal 22 amino acids (residues 267–288) can function in *trans* to the active core within a mixed multimer. However, a mixture of $IN^{1-266/D116N}$ and IN^{50-288} was also active (Figure 6A and B, lanes 5 and 6), demonstrating that the C-terminus can also act efficiently in *cis* to the core that is competent for polynucleotidyl transfer. We conclude that the C-terminal 22 amino acid residues can function either in *cis* or *trans* to the active core in the mixed multimers.

Amino acid residues 187–234 are important in cis to the active core for efficient 3′ processing and DNA strand transfer

We have shown that the core alone (amino acid residues 50–186) is sufficient for disintegration activity (Bushman *et al.*, 1993). The results presented below show that additional sequences, between this domain and the C-terminal region, are important in *cis* to the active core for efficient 3′ processing and DNA strand transfer in the complementation assays reported here.

Mixtures of either of three deletion derivatives that lacked both the N- and C-termini (IN^{50-263}, IN^{50-254}, or IN^{50-234}) together with D116N gave similar levels of 3′ processing and strand transfer activities as mixtures of either IN^{1-266} or IN^{50-288} with D116N (Figure 7A and B, lanes 3 and 5–8). Evidently, the D116N protein can supply both the N- and C-terminal functions required for 3′ processing and strand transfer in *trans* to an active core for these three deletion mutants. However, reactions containing either of the more extensive C-terminal deletion mutants IN^{50-212} or IN^{50-186}, together with D116N, yielded ~10-fold less product than IN^{50-234} and D116N (Figure 7A and B, lanes 8–10). Therefore, a region between amino acid residues 187 and 234 is important in *cis* to the active core in assays of 3′ processing and DNA strand transfer, but not disintegration.

Discussion

Certain mutant HIV-1 integrase proteins containing extensive deletions or point mutations that render them inactive for 3′ processing and DNA strand transfer exhibit robust activity when assayed as mixtures. We know of no other example of a DNA recombination reaction in which mutant recombination proteins, each deleted for essential functional regions, complement to restore essentially wild type activity. This result provided a tool with which to probe the organization of a functional integrase multimer. We find that HIV-1 integrase comprises several discrete functional domains, some of which may be contributed by separate monomers within the active multimer. Our findings closely parallel those reported in this issue by van Gent *et al.* (1993).

HIV-1 integrase functions as a multimer in the 3′ processing and DNA strand transfer reactions

The complementation *in vitro* among mutant HIV-1 integrase proteins that displayed little or no activity alone implies that 3′ processing and DNA strand transfer are catalyzed by a multimer of integrase. Recent studies indicate that the integrase protein of Rous sarcoma virus also acts as a multimer in assays for 3′ processing and strand transfer *in vitro* (Jones *et al.*, 1992).

Mixing of two deletion derivatives of HIV-1 integrase that share only the core domain (IN^{1-173} and IN^{50-288}, see Figure 3) resulted in a significant increase in 3′ processing and strand transfer activity compared with reactions containing either protein alone. Complementation by these two mutant proteins cannot involve contacts between one N-terminal region and another, or between one C-terminal region and another. It is likely that the core domain carries the information necessary for multimerization, although the possibility of interactions involving non-equivalent parts of integrase is not excluded.

The results of protein–protein cross-linking indicate that the core domain is capable of multimerization. Cross-linking of wild type integrase in solution revealed a prominent dimeric species and a less abundant tetrameric form (Figure 4A). Dimers were also the most prominent cross-linked products of deletion derivatives of integrase that contain only the core domain (Figure 4B). Consistent with

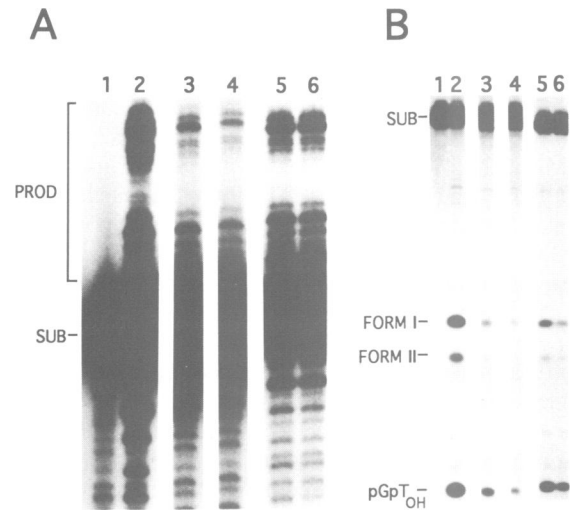

Fig. 6. The C-terminal region can function in *cis* or *trans* to the active core. (**A**) Denaturing polyacrylamide gel of DNA strand transfer products. Integrase was omitted from the reaction in lane 1; lane 2 contained wild type integrase; lane 3, mixture of D116N and IN^{1-266}; lane 4, mixture of E152Q and IN^{1-266}; lane 5, mixture of IN^{1-266} and IN^{50-288}; lane 6, mixture of $IN^{1-266/D116N}$ and IN^{50-288}. The reactions in lanes 5 and 6 were from a separate polyacrylamide gel. The relative activities of these protein mixtures are shown in Figure 3. Other labeling is the same as in Figure 2A. (**B**) Denaturing polyacrylamide gel of 3′ processing products. The reactions contained the same integrase proteins as in panel A. The reactions in lanes 5 and 6 were from a separate polyacrylamide gel. Other labeling is the same as in Figure 2B.

the view that this multimerization is functionally relevant, the core alone is sufficient for disintegration activity (Bushman *et al.*, 1993) and the results of preliminary experiments suggest that disintegration may also be mediated by a multimer of integrase; HIV-1 integrase mutants D64N and IN^{1-173}, which have little and no activity alone, respectively, display a significantly elevated level of disintegration activity upon mixing (A.Engelman and R.Craigie, unpublished data)

Is there a distinct motif in the core domain that is important for multimerization? The conserved residues of the D,D-35-E motif, which are essential for polynucleotidyl transfer, do not appear to be important for integrase–integrase recognition. Mixtures of IN^{50-288} and either of three mutants containing a single substitution of one of the conserved residues (D116N, E152Q or D64N) were active in assays for 3′ processing and DNA strand transfer (Figure 5; A.Engelman and R.Craigie, unpublished data). Mixtures of IN^{1-173} and IN^{50-288} were equally active regardless of whether the core of the IN^{1-173} partner contained the D116N amino acid substitution (Figure 3). Furthermore, analysis of D116N by cross-linking revealed a pattern of cross-linked species indistinguishable from that produced by wild type integrase (data not shown).

A potential leucine zipper motif, located near the C-terminus of the core domain, has been suggested to be involved in multimerization (Lin and Grandgenett, 1991). However, this hypothesis has not yet been tested.

Organization of the functional multimer

By comparing complementation among different combinations of mutant proteins, we were able to investigate the domain organization within active multimers of integrase. Although the core domain by itself displays disintegration

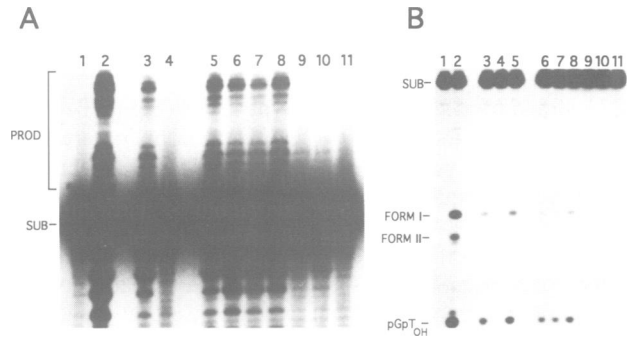

Fig. 7. Amino acid residues 187–234 are important in *cis* to the active core. (**A**) Denaturing polyacrylamide gel of DNA strand transfer products. Integrase was omitted from the reaction in lane 1; lane 2 contained wild type integrase; lane 3, mixture of D116N and IN^{1-266}; lane 4, mixture of D116N and IN^{1-173}; lane 5, mixture of D116N and IN^{50-288}; lane 6, mixture of D116N and IN^{50-263}; lane 7, mixture of D116N and IN^{50-254}; lane 8, mixture of D116N and IN^{50-234}; lane 9, mixture of D116N and IN^{50-212}; lane 10, mixture of D116N and IN^{50-186}; lane 11, mixture of D116N and IN^{50-173}. Other labeling is the same as in Figure 2A. (**B**) Denaturing polyacrylamide gel of 3′ processing reactions. The products contained the same integrase proteins as in panel A. Other labeling is the same as in Figure 2B.

activity (Bushman *et al.*, 1993), the present results indicate that three additional regions of integrase are important for 3′ processing and DNA strand transfer.

We found that mixtures of two mutant proteins result in efficient 3′ processing and strand transfer only if the N-terminal region of integrase and the core domain that is capable of polynucleotidyl transfer are present on the different integrase monomers. We infer that these two domains also work in *trans* in multimers containing only wild type integrase.

The C-terminal region of integrase functioned either in *trans* or *cis* to the active core in the mixed multimers. This result is somewhat surprising since it suggests that this region can function in non-equivalent positions within the multimer. It is possible that this region is sufficiently flexible to assume essentially the same position when present in *cis* or *trans* to the active core. Alternatively, the C-terminal region may be able to function at more than one distinct position within the multimer; presumably such positions would be symmetrically related.

A region between amino acid residues 187 and 234 was found to be important in *cis* to the active core in assays of 3′ processing and DNA strand transfer. This conclusion is based on the activities of mixtures of D116N and a set of disintegration-competent mutants containing deletions extending from the C-terminus. Mixtures of either IN^{50-212} or IN^{50-186} with D116N displayed only minuscule levels of activities, in contrast to the robust activities of the mixture of IN^{50-234} and D116N (Figure 7). These results identify a functional boundary between residues 213 and 234; sequences to the N-terminal side of this boundary must be present in *cis* to the active core for efficient 3′ processing and DNA strand transfer, whereas the functional domain on the C-terminal side of this boundary may be present either in *cis* or *trans*.

Apparent redundancy of functional domains in active multimers of HIV-1 integrase

The experiments presented here reveal a redundancy of functional domains within the active integrase multimers that

mediate the *in vitro* reactions. For example, the mixture of $IN^{1-173/D116N}$ and IN^{50-288} lacks one or more functional domains on each partner, yet promotes 3′ processing and DNA strand transfer at near wild type levels (Figure 3). We speculate that although some functional domains are not required on both partners in our assays, a full complement of domains is required for the pairwise integration of two viral DNA ends. Whereas integration *in vivo* requires cleavage and insertion of pairs of viral DNA ends into both strands of a target DNA, the *in vitro* assays used here do not distinguish single end from two end insertion products. In fact, assays modified to discriminate between these two types of product indicate that essentially all the insertions mediated by purified HIV-1 integrase are 'half reactions', resulting from insertion of a single viral DNA end into a single strand of the target DNA (Bushman and Craigie, 1991). This 'half-reaction' may require only a 'half-set' of the functional domains within the multimer that mediates pairwise insertion.

What particular functions are carried out by the individual protein domains? Although the core domain of HIV-1 integrase has been demonstrated to catalyze polynucleotidyl transfer (Bushman *et al.*, 1993), the biochemical functions of the other domains have not yet been rigorously established. It has been suggested that the N-terminal domain may be important for sequence-specific interactions with the viral DNA ends, as mutations within the domain preferentially affect 3′ processing and strand transfer but not disintegration activity *in vitro* (Engelman and Craigie, 1992; van Gent *et al.*, 1992; Bushman *et al.*, 1993; Leavitt *et al.*, 1993; Vincent *et al.*, 1993). The conserved His and Cys residues within this domain may form part of a folded structure that includes a bound metal ion; a peptide comprising the N-terminal region folds stably in the presence of zinc (Burke *et al.*, 1992) and the conserved His and Cys residues are important for zinc binding by the intact protein *in vitro* (Bushman *et al.*, 1993). Sequences to the C-terminal side of the core may contribute to interactions with target DNA, as deletions of these regions of HIV-1 integrase affect the non-specific DNA binding activity of the protein (van Gent *et al.*, 1991; Schauer and Billich, 1992; Woerner *et al.*, 1992).

If the N-terminal domain of HIV-1 integrase is important for sequence-specific interactions with the viral DNA ends, the results presented here imply that different monomers within an active multimeric complex are responsible for viral DNA binding and catalysis of polynucleotidyl transfer. An intimate division of labor between 'identical' polypeptides within a multimeric complex is not without precedent in DNA recombination reactions. Within an active tetramer of Flp recombinase, a DNA site bound by one monomer is cleaved in *trans* by a tyrosine residue from a separate monomer (Chen *et al.*, 1992).

Implications for integration in vivo

We recognize that the complexes made between integrase and DNA substrate *in vitro* may not fully reflect the active species *in vivo*, but these complexes can be expected to reveal the minimal protein–protein and protein–DNA interactions required for the biochemical steps of retroviral DNA integration: 3′ processing of the viral DNA ends and the DNA strand transfer reaction that inserts these ends into a target DNA. Additional interactions are likely to be required for the full integration reaction *in vivo*.

Viral DNA made by reverse transcription *in vivo* forms part of a high molecular weight preintegration complex that is derived from the viral core (Brown *et al.*, 1987; Fujiwara and Mizuuchi, 1988; Bowerman *et al.*, 1989; Ellison *et al.*, 1990; Farnet and Haseltine, 1990, 1991). This nucleoprotein complex, isolated from infected cells, efficiently integrates the DNA *in vitro*; essentially all the integration products appear to result from pairwise insertion of the viral DNA ends. Why are such complexes not efficiently formed upon mixing integrase with DNA substrates? Since retroviruses function as nucleoprotein complexes that may contain all the machinery needed for each successive step of the replication cycle leading to integration, recapitulation of part of this cycle may be required to form preintegration complexes *in vitro* with the properties of complexes made *in vivo*. It is also possible that efficient formation of such higher order complexes simply requires additional protein factors, either as an integral part of the structure or as a catalyst for complex formation. The phage Mu DNA transposition reaction, which proceeds by essentially identical chemical steps as retroviral DNA integration (Mizuuchi, 1992), offers such a precedent. With a linear DNA substrate, Mu transposase exhibits only specific binding to its recognition sequences and this binding, like the binding of purified integrase to DNA substrates (van Gent *et al.*, 1991), is efficiently competed by challenge with competitor DNA (Mizuuchi *et al.*, 1991). However, when the DNA is supercoiled and an accessory protein together with metal ion is also present, a highly stable nucleoprotein complex is formed that is resistant to subsequent challenge by competitor DNA (Mizuuchi *et al.*, 1991, 1992). The accessory factors presumably lower an activation energy barrier, resulting in a much faster rate of formation of the stable reaction intermediate. A major task in furthering our understanding of retroviral DNA integration is to discover how to reconstitute efficiently a stable preintegration complex of viral DNA substrate, integrase and any required accessory factors that possesses the properties of such complexes isolated from infected cells.

Materials and methods

Plasmids

Plasmids encoding wild type integrase, the point mutants D116N and E152Q (Engelman and Craigie, 1992) and the deletion mutants IN^{50-263}, IN^{50-212}, IN^{50-186} and IN^{50-173} (Bushman *et al.*, 1993) have been described. Plasmids encoding the other deletion derivatives used in this study were prepared essentially as described by Bushman *et al.* (1993).

Plasmids encoding deletion mutants with the D116N amino acid substitution were prepared as follows. For $IN^{50-288/D116N}$, the plasmid encoding IN^{50-288} was cut with *Nsi*I and *Bam*HI and the 5.7 kb DNA fragment containing the vector and 5′-proximal sequences of integrase was isolated. Plasmid pIND116N, which encodes D116N, was also cut with *Nsi*I and *Bam*HI and the 0.72 kb fragment containing 3′-proximal sequences of integrase, including the codon for the D116N substitution, was isolated. The plasmid encoding $IN^{50-288/D116N}$ was produced by ligation of these two DNA fragments. For $IN^{1-173/D116N}$ and $IN^{1-266/D116N}$, plasmids encoding the deletion derivatives were cut with *Nde*I and *Afl*II and the corresponding DNA fragments, which contained the vector and 3′-proximal sequences of integrase, were isolated. Plasmid pIND116N was also cut with *Nde*I and *Afl*II and the 0.52 kb fragment containing 5′-proximal sequences of integrase with the point mutation was isolated. Plasmids encoding each double mutant were produced by ligation of the appropriate DNA fragments. The presence of the base substitution in each of the three plasmids encoding the double mutants was confirmed by dideoxy sequencing.

Proteins

All deletion derivatives contained a 20 amino acid residue 'His-Tag' motif at the N-terminus of the proteins to facilitate purification. Deletion mutants were purified from insoluble extracts of *Escherichia coli* cells induced for expression essentially as described (Bushman *et al.*, 1993). This procedure included two passes over a Ni^{2+}-charged metal-chelating column in the presence of 6 M guanidine–HCl and removal of the denaturant by stepwise dialyses at 4°C. The final dialysis buffer (CN buffer) contained 200 mM NaCl, 20 mM HEPES (pH 7.5), 1 mM DTT, 0.1 mM EDTA, 10% glycerol (w/v) and 15 mM CHAPS. Wild type integrase and the point mutants D116N and E152Q were purified as described by Engelman and Craigie (1992). These proteins were also dialyzed against CN buffer at 4°C.

In addition to the N-terminal His-Tag motif, IN^{1-173} and IN^{1-266} contained a 10 amino acid motif recognized by monoclonal antibody 9E10 (Kolodziej and Young, 1991) between the His-Tag motif and the first amino acid of integrase. Deletion derivatives IN^{50-288} and $IN^{50-288/D116N}$ contained a nine amino acid motif recognized by monoclonal antibody 12CA5 (Kolodziej and Young, 1991) between the His-Tag motif and amino acid residue 50 of integrase. The *in vitro* activities of purified deletion derivatives were identical regardless of the presence of these additional amino acid motifs, both when assayed alone for disintegration or as mixtures for 3′ processing and DNA strand transfer (data not shown).

DNA strand transfer, 3′ processing and disintegration reactions

DNA strand transfer, 3′ processing and disintegration reactions were incubated for 1 h at 37°C and included 25 mM MOPS (pH 7.2), 10 mM 2-mercaptoethanol, 100 μg/ml BSA, 7.5 mM $MnCl_2$, 10% glycerol (w/v), 17 mM NaCl, 0.75 mM CHAPS, 25 nM DNA substrate and 160 nM integrase in a volume of 16 μl. Substrate DNAs were prepared as previously described by Engelman and Craigie (1992). For mixing experiments, proteins were premixed at an equimolar ratio in CN buffer on ice prior to addition to reactions to a final total integrase protein concentration of 160 nM. Strand transfer, 3′ processing and disintegration reactions were prepared and reactions were stopped and portions were electrophoresed in 20% polyacrylamide denaturing gels, as previously described by Engelman and Craigie (1992). Quantitation of reaction products was done with a PhosphorImager (Molecular Dynamics).

Protein–protein cross-linking

Wild type or mutant integrase was treated with 40 μg/ml disuccinimidyl suberate (Pierce) unless otherwise noted. Reactions included 45 μg/ml integrase in 10 μl of buffer containing 20 mM HEPES (pH 7.5), 60 mM NaCl, 0.7 mM EDTA, 10% (w/v) glycerol and 4.5 mM CHAPS. Following 10 min at 22°C, reactions were quenched by the addition of lysine and Tris–HCl (pH 8.0) to the final concentrations of 3 mM and 25 mM, respectively. After 10 min at 22°C, reactions were boiled for 10 min in protein sample buffer [50 mM Tris (pH 6.8), 1.5% SDS (w/v), 7.5% glycerol (w/v), 4% 2-mercaptoethanol (w/v), 50 mM DTT and 0.001% (w/v) bromophenol blue] and electrophoresed in Tris–glycine SDS–polyacrylamide gels (Novex). Gels were stained for 15 min with Coomassie blue G-250 followed by destaining with 5% (v/v) methanol, 7.5% (v/v) acetic acid. Destained gels were treated with 50% (v/v) methanol overnight, followed by staining with $AgNO_3$ as described by Wray *et al.* (1981).

Acknowledgements

We thank Martin Gellert, Kiyoshi Mizuuchi, Howard Nash, David Roth and Harri Savilahti for careful reading of the manuscript. This work was supported in part by the National Institutes of Health Intramural AIDS Targeted Antiviral Program. F.D.B. is a Special Fellow of the Leukemia Society of America.

References

Bowerman,B., Brown,P.O., Bishop,J.M. and Varmus,H.E. (1989) *Genes Dev.*, **3**, 469–478.

Brown,P.O., Bowerman,B., Varmus,H.E. and Bishop,J.M. (1987) *Cell*, **49**, 347–356.

Burke,C.J., Sanyal,G., Bruner,M.W., Ryan,J.A., Lafemina,R.L., Robbins,H.L., Zeft,A.S., Middaugh,C.R. and Cordingley,M.G. (1992) *J. Biol. Chem.*, **267**, 9639–9644.

Bushman,F.D. and Craigie,R. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 1339–1343.

Bushman,F.D., Fujiwara,T. and Craigie,R. (1990) *Science*, **249**, 1555–1558.

Bushman,F.D., Engelman,A., Palmer,I., Wingfield,P. and Craigie,R. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 3428–3432.

Chen,J.-W., Lee,J. and Jayaram,M. (1992) *Cell*, **69**, 647–658.

Chow,S.A., Vincent,K.A., Ellison,V. and Brown,P.O. (1992) *Science*, **255**, 723–726.

Drelich,M., Wilhelm,R. and Mous,J. (1992) *Virology*, **188**, 459–468.

Ellison,V., Abrams,H., Roe,T., Lifson,J. and Brown,P. (1990) *J. Virol.*, **64**, 2711–2715.

Engelman,A. and Craigie,R. (1992) *J. Virol.*, **66**, 6361–6369.

Engelman,A., Mizuuchi,K. and Craigie,R. (1991) *Cell*, **67**, 1211–1221.

Farnet,C. and Haseltine,W.A. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 4164–4168.

Farnet,C. and Haseltine,W.A. (1991) *J. Virol.*, **65**, 1910–1915.

Fujiwara,T. and Mizuuchi,K. (1988) *Cell*, **54**, 497–504.

Goff,S.P. (1992) *Annu. Rev. Genet.*, **26**, 527–544.

Johnson,M.S., McClure,M.A., Feng,D.-F., Gray,J. and Doolittle,R.F. (1986) *Proc. Natl Acad. Sci. USA*, **83**, 7648–7652.

Jones,K.S., Coleman,J., Merkel,G.W., Laue,T.M. and Skalka,A.M. (1992) *J. Biol. Chem.*, **267**, 16037–16040.

Khan,E., Mack,J.P.G., Katz,R.A., Kulkosky,J. and Skalka,A.M. (1991) *Nucleic Acids Res.*, **19**, 851–860.

Kolodziej,P.A. and Young,R.A. (1991) *Methods Enzymol.*, **194**, 508–519.

Kulkosky,J., Jones,K.S., Katz,R.A., Mack,J.P.G. and Skalka,A.M. (1992) *Mol. Cell. Biol.*, **12**, 2331–2338.

Lafemina,R.L., Callahan,P.L. and Cordingley,M.G. (1991) *J. Virol.*, **65**, 5624–5630.

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.

Leavitt,A.D., Shiue,L. and Varmus,H.E. (1993) *J. Biol. Chem.*, **268**, 2113–2119.

Lin,T.-H. and Grandgenett,D.P. (1991) *Protein Engng*, **4**, 435–441.

Mizuuchi,K. (1992) *Annu. Rev. Biochem.*, **61**, 1011–1051.

Mizuuchi,M., Baker,T.A. and Mizuuchi,K. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 9031–9035.

Mizuuchi,M., Baker,T.A. and Mizuuchi,K. (1992) *Cell*, **70**, 303–311.

Schauer,M. and Billich,A. (1992) *Biochem. Biophys. Res. Commun.*, **185**, 874–880.

Sherman,P.A. and Fyfe,J.A. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 5119–5123.

van Gent,D.C., Elgersma,Y., Bolk,M.W.J., Vink,C. and Plasterk,R.H.A. (1991) *Nucleic Acids Res.*, **19**, 3821–3827.

van Gent,D.C., Oude Groeneger,A.A.M. and Plasterk,R.H.A. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 9598–9602.

van Gent,D.C., Vink,C., Oude Groeneger,A.A.M. and Plasterk,R.H.A. (1993) *EMBO J.*, **12**, 3261–3267.

Vincent,K.A., Ellison,V., Chow,S.A. and Brown,P.O. (1993) *J. Virol.*, **67**, 425–437.

Vink,C., van Gent,D.C., Elgersma,Y. and Plasterk,R.H.A. (1991a) *J. Virol.*, **65**, 4636–4644.

Vink,C., Yeheskiely,E., van der Marel,G.A., van Boom,J.H. and Plasterk,R.H.A. (1991b) *Nucleic Acids Res.*, **19**, 6691–6698.

Whitcomb,J.M. and Hughes,S.H. (1992) *Annu. Rev. Cell Biol.*, **8**, 275–306.

Woerner,A.M., Klutch,M., Levin,J.G. and Marcus-Sekura,C.J. (1992) *AIDS Res. Human Retrovir.*, **8**, 297–304.

Wray,W., Boulikas,T., Wray,V.P. and Hancock,R. (1981) *Anal. Biochem.*, **118**, 197–203.

Received on March 26, 1993; revised on May 10, 1993