The GP(Y/F) Domain of TF1 Integrase Multimerizes when Present in a Fragment, and Substitutions in This Domain Reduce Enzymatic Activity of the Full-length Protein

Hirotaka Ebina, Atreyi Ghatak Chatterjee, Robert L. Judson, and Henry L. Levin

From the Section on Eukaryotic Transposable Elements, Laboratory of Gene Regulation and Development, NICHD, National Institutes of Health, Bethesda, Maryland 20892

Integrase (INs) of retroviruses and long terminal repeat retrotransposons possess a C-terminal domain with DNA binding activity. Other than this binding activity, little is known about how the C-terminal domain contributes to integration. A stretch of conserved amino acids called the GP(Y/F) domain has been identified within the C-terminal IN domains of two distantly related families, the γ-retroviruses and the metavirus retrotransposons. To enhance understanding of the C-terminal domain, we examined the function of the GP(Y/F) domain in the IN of Tf1, a long terminal repeat retrotransposon of Schizosaccharomyces pombe. The activities of recombinant IN were measured with an assay that modeled the reverse of integration called disintegration. Although deletion of the entire C-terminal domain disrupted disintegration activity, an alanine substitution (P365A) in a conserved amino acid of the GP(Y/F) domain did not significantly reduce disintegration. When assayed for the ability to join two molecules of DNA in a reaction that modeled forward integration, the P365A substitution disrupted activity. UV cross-linking experiments detected DNA binding activity in the C-terminal domain and found that this activity was not reduced by substitutions in two conserved amino acids of the GP(Y/F) domain, G364A and P365A. Gel filtration and cross-linking of a 71-amino acid fragment containing the GP(Y/F) domain revealed a surprising ability to form dimers, trimers, and tetramers that was disrupted by the G364A and P365A substitutions. These results suggest that the GP(Y/F) residues may play roles in promoting multimerization and intermolecular strand joining.

Retroviruses and long terminal repeat (LTR)3 retrotransposons are closely related elements that depend on integrase (IN) to insert their cDNA into the genome of host cells. IN proteins are composed of three structurally distinct domains (1). The N-terminal domains contain an HHCC motif, and the catalytic core domains in the center of INs possess the DDE residues that mediate catalysis. The C-terminal domains of INs have little sequence conservation but possess nonspecific DNA binding activity.

INs expressed as recombinant proteins possess a variety of catalytic activities. Oligonucleotides that model the termini of LTRs are trimmed by IN in a processing reaction that removes terminal nucleotides 3’ of the conserved CA. Once the processing reaction is complete, strand transfer occurs. In this reaction, the 3’-hydroxyls of the terminal A serve as nucleophiles in transesterification reactions that cleave phosphodiester bonds in the target DNA and make covalent bonds between the 3’-ends of the viral DNA and the 5’-ends of the target DNA (2, 3). Under highly specific conditions, IN can catalyze concerted integration, the simultaneous insertion of two ends of donor DNA into the same site of target DNA (4–7). In addition, INs also catalyze disintegration, the reverse of integration that uses model substrates that mimic one end of an LTR inserted into target DNA (8, 9). This assay is a particularly sensitive method for measuring catalytic activity of INs, perhaps because it detects strand breaking and joining within a single substrate molecule, an intramolecular reaction.

Currently, no molecular structures exist of an IN that possesses all three domains. The structures that do exist are of individual domains and do not include bound DNA. As a result, the function of the C-terminal domain in integration is not clear. An additional difficulty in determining the function of the C-terminal domains stems from their low levels of sequence conservation. However, close examination of C termini did identify two separate sequence modules that exist either alone or in combination in a wide variety of INs (10). One module termed the GP(Y/F) domain is present in the INs of a diverse set of LTR retrotransposons of the Metaviridae family (formally called the Ty3/gypsy family) (10, 11). In addition, the GP(Y/F) domain is also present in the distantly related genus of γ-retroviruses that includes the Moloney murine leukemia virus (M-MuLV). The function of the GP(Y/F) domain has not been studied. The chromodomain (CHD) is another module disco-

3-(3-cholamidopropyl)dimethylammonium]-1-propanesulfonic acid; Bis- Tris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol; aa, amino acid(s); HIV-1, human immunodeficiency virus, type 1.
The GP(Y/F) Domain of IN

ered in the C termini of INs and is present in the metavirus genus of LTR retrotransposons. The CHD is similar to the domains in HP1 proteins that mediate the formation of heterochromatin by binding histone H3 when methylated at Lys-9. Recent results indicate that CHDs in some INs do have interactions with histone H3 methylated at Lys-9 (12).

Tf1 is an LTR retrotransposon of Schizosaccharomyces pombe that integrates specifically upstream of polymerase II-transcribed genes (13–15). The IN of Tf1 possesses the HHCC motif near the N terminus and the DDE motif in the central region. Interestingly, the C-terminal portion of the Tf1 IN possesses both the GP(Y/F) domain and the CHD (10). Recent experiments revealed that Tf1 IN purified as a recombinant protein possesses significant activity in assays that measure 3’ processing, strand transfer, and disintegration (16). Assays of Tf1 IN without the CHD revealed the surprising result that the CHD restricts catalytic activity by as much as 8-fold (16).

The experiments reported here use the IN of Tf1 as a model in order to study the function of the GP(Y/F) domain. A series of deletions in recombinant IN revealed that the C-terminal domain was required for disintegration activity. However, a single amino acid substitution in a conserved amino acid of the GP(Y/F) domain (P365A) did not significantly reduce disintegration. Assays for strand transfer activity revealed the P365A substitution significantly reduced activity. The results of gel filtration and chemical cross-linking indicated that a 71-aa fragment containing the GP(Y/F) domain formed dimers, trimers, and tetramers. Single amino acid substitutions in conserved residues of the GP(Y/F) domain, G364A and P365A, abrogated this multimerization. These data suggest that the GP(Y/F) residues may promote multimerization and strand transfer activity.

**EXPERIMENTAL PROCEDURES**

**Plasmids**—To generate versions of Tf1 with substitutions in the GP(Y/F) domain, fusion-PCR fragments with mutations were inserted into the NdeI and BamHI sites of plH414-2 (Wt Tf1-neo). The primers are described in Table S1. Transposition assays were performed as previously described (17).

The construction of plH2468, the plasmid for the expression of the full-length Tf1 IN and the plasmid expressing IN lacking the CHD (CH−), plH2469, have been described previously (16). The plasmids expressing the fragments of IN were generated by similar methods. The DNA fragments coding for each protein were amplified by PCR with the Pfu Ultra Hotstart 2× Master Mix (Stratagene) and primer pairs as indicated in the supplemental data (Table S1). The DNA generated was cleaved with NdeI and BamHI and cloned into the vector pET15b cut with NdeI and BamHI. Each insert was sequenced. All plasmids are listed in Table S2.

The Purification of His-tagged Recombinant Proteins—BL21 cells containing the expression plasmids were grown at 32 °C until they reached an A600 of ~0.6. Inductions were performed at 16 °C with 1 mM isopropyl 1-thio-β-D-galactopyranoside for 24 h. The cells were harvested, and pellets were stored at −80 °C.

The methods of protein purification were based on our previous report but contained the following modifications (16). The volumes of the cultures were 250 ml, and the sonication was performed in 30 ml of buffer. The bed volumes of Sepharose-6B (BD Talon, BD Biosciences) for the columns were 2.0 ml. All column treatments were by gravity flow. The column washes were 50 ml with no imidazole followed by 100 ml with 25 mM imidazole. The proteins were eluted with a sequence of 4-ml steps that contained 35, 50, 75, 100, and 150 mM imidazole.

The fractions eluted from the column were dialyzed into a storage buffer (50 mM Na2HPO4, pH 7.5, 10% glycerol, 1 mM EDTA, 1 mM dithiothreitol, and 0.5 mM NaCl) and frozen at −80 °C.

**Purification of IN Lacking the His Tag and Partial Trypsin Digestion**—This purification was based on the previous report with modifications that differed from the purification of the His tag-containing IN described above (16). Harvested cells with plH2468 were incubated for 15 min on ice in 50 mM HEPES, pH 7.5, buffer, 0.1 mM NaCl, 1 mM phenylmethylsulfonyl fluoride, 2 mM β-mercaptoethanol, 0.5 mg/ml lysozyme, and 1× Complete EDTA-free protease inhibitor mixture (Roche Applied Science). The cells were sonicated, CHAPS was added to the lysate to a final concentration of 0.05%, and the sample was inverted six times. The sample was then spun in an SW28 at 25,000 rpm for 1.5 h. After ultracentrifugation, the supernatant was loaded onto a 5-ml HIS Trap FF column (Amersham Biosciences). At 1 ml/min, the column was washed with the lysis buffer plus the CHAPS and 0.5 mM NaCl and 20 mM imidazole. The protein was eluted into 2-ml fractions with an imidazole gradient of 40–400 mM.

IN-containing fractions were dialyzed into 75 mM NaCl with 50 mM HEPES-NaOH (pH 7.5), 10 mM MgSO4, 10% glycerol, 1 mM EDTA, 2 mM dithiothreitol, and 0.1% CHAPS and loaded onto a 5-ml heparin-Sepharose FF column (Amersham Biosciences). The protein was eluted with 0.5 mM NaCl. IN-containing fractions were combined and dialyzed into thrombin (Sigma)-containing buffer to remove the His tag. Thrombin was removed with a benzamidine column (Amersham Biosciences). The IN was stored in 50 mM HEPES, pH 7.5, buffer, 0.5 mM NaCl, 1 mM EDTA, 2 mM dithiothreitol, 10% glycerol, and 0.1% CHAPS at a final concentration of 0.73 μg/μl.

Either 50 or 500 ng of trypsin (Roche Applied Science) was added to 18 μg of Tf1 IN in the above storage buffer and incubated at 30 °C. Reactions were stopped and collected at 0.5, 2, 4.5, and 21.5 h by the addition of 2 mM phenylmethylsulfonyl fluoride. A mock reaction, containing no trypsin, was incubated for 21.5 h at 30 °C. Reactions were divided and loaded on both 20% Tris-glycine SDS-polyacrylamide gels (Invitrogen) for band visualization and 4–12% NuPAGE BisTris gels (Invitrogen) for N terminus analysis. Tris-glycine gels were stained with Coomassie Brilliant Blue R250. The protein from the NuPAGE gels was electrotransferred to the Immobilon P membrane (Millipore). The membrane was Coomassie-stained, and bands were cut out and sent for N terminus sequencing at the Food and Drug Administration Center for Biologics Evaluation and Research by Dr. Ngo Nguyen.

**Disintegration and Strand Transfer Assays**—The disintegration and strand transfer assays were conducted as described (16). However, in this work, the reactions were stopped by add-
ing 10 μl of loading buffer (50% glycerol, 250 mM EDTA, 0.5 mg/ml bromphenol blue, and 0.05 mg/ml xylene cyanol) and heating at 95 °C for 3 min. The samples were loaded on an 8 M urea, 14% (w/v) polyacrylamide sequencing gel and electrophoresed in Tris-borate EDTA buffer.

HL1127 and HL1034 were the DNA oligonucleotide that were 5’-end-labeled for use in the disintegration and strand transfer assays, respectively (Table S1) (16). Aliquots containing 0.4 μg of each oligonucleotide were 32P-end-labeled using 20 units of T4 polynucleotide kinase (New England Biolabs) and 100 μCi of [γ-32P]ATP in a volume of 20 μl of the buffer from the supplier. After an incubation for 1 h at 37 °C, 1 μl of 0.25 M EDTA was added to each sample, and the mixtures were treated for 5 min at 95 °C to inactivate the enzyme. After adding NaCl to a final concentration of 0.1 M and increasing the volume to 40 μl, the labeled DNA was annealed to 3 times the molar ratio of the appropriate nonlabeled DNA. After 5 min at 95 °C, the mixtures were allowed to cool to room temperature for about 2 h. The duplexes generated were then stored at −20 °C and were used after freezing and thawing for up to 1 month. One pmol of substrate was added to each reaction.

**Pull-down Experiment**—Binding of His-tagged INs to full-length IN lacking a His tag was assayed in a binding buffer containing 50 mM HEPES (pH 7.5), 0.5 M NaCl, 0.1% Triton X-100. Six micromolars of recombinant His-tagged IN was incubated with 6 μg of full-length IN lacking the tag in 500 μl of binding buffer. Following a 30-min incubation at 4 °C, the reactions were supplemented with 40 μl of prewashed Ni2+–nitrilotriacetic acid–agarose (Qiagen) and stirred for an additional 60 min at 4 °C. The agarose beads were recovered by centrifugation for 1 min at 960 × g at 4 °C and washed with 500 μl of binding buffer two times and then washed three times with 500 μl of binding buffer supplemented with 25 mM imidazole. Bound proteins were eluted in 40 μl of binding buffer supplemented with 400 mM imidazole and analyzed on a 10–20% SDS-polyacrylamide gel. The proteins in the gel were transferred to Immobilon-P membranes (Millipore). The membrane was probed with anti-IN rabbit antibody (1:10,000) (41). The secondary antibody was horseradish peroxidase-conjugated donkey anti-rabbit Ig, whole antibody (1:10,000; Amersham Biosciences). ECL Plus was used to detect the protein signals (Amersham Biosciences).

**Size Exclusion Chromatography**—Size exclusion chromatography was performed on a Superdex 75 10/300 GL column or Superdex 200 PC 3.2/30 column on an AKTA FPLC system (GE Healthcare). A flow rate of 0.5 ml/min with a mobile phase of 50 mM HEPES, pH 7.5, 0.5 M NaCl, 1% (v/v) glycerol, 1 mM EDTA, 1 mM dithiothreitol was used for the Superdex 75HR 10/300 GL column. A flow rate of 0.25 ml/min was used for the Superdex 200 PC 3.2/30 column. Typically, 100 μg of purified IN polypeptides at a concentration of 1 μg/μl were injected. The concentrations of the proteins and fragments were as follows: IN, 17.4 μM; core, 32.7 μM; ΔC-(1–334), 24.3 μM; GP(Y/F) fragment, 94.3 μM; CHD, 105.3 μM. Samples were subjected to centrifugation for 5 min at 10,000 × g prior to injection on the column. Absorbance of the column eluate was monitored at 280 nm. Samples from peak fractions were monitored by SDS-PAGE for the presence of the expected protein species. The column was calibrated using five different globular proteins as molecular weight standards (Gel Filtration Calibration Kits, High Molecular Weight and Low Molecular Weight; Amersham Biosciences), and the apparent molecular weight of each sample peak was determined using linear regression of the log of known molecular weight versus the elution behavior (Kav or elution time).

**Chemical Cross-linking**—Cross-linking of the GP(Y/F) fragment was conducted using bis(sulfosuccinimidyl) suberate (Pierce) at concentrations of 0.2, 1.0, and 2.0 mM in reactions (100 mM HEPES, pH 7.5, 0.5 M NaCl) that were incubated at room temperature for 60 min. The GP(Y/F) fragment was at a concentration of 25 μM. The reactions were quenched by adding Tris-HCl to a final concentration of 60 mM and incubating at room temperature for 15 min. One-half volume of 2× sample buffer without 2-mercaptoethanol (125 mM Tris-HCl, 5% SDS, 20% glycerol, 0.5 mg/ml bromphenol blue) was added to the reactions, and the samples were heated at 95 °C for 10 min. Covalently linked multimers were detected by separation in 10–20% SDS-polyacrylamide gels and silver staining.

**DNA Binding Assay**—Reaction mixtures for UV cross-linking contained the same buffer components as the disintegration reactions, except the concentration of glycerol was 1% (v/v). Unless otherwise stated, 1 μg of protein was combined with 1 pmol of DNA per reaction. The following were the molar quantities of protein added to 1 pmol of DNA: IN, 17.4 pmol; CHD, 105.3 pmol; N-terminal domain (NTD), 66.2 pmol; GP(Y/F), 94.3 pmol; CHD, 105.3 pmol. The mixtures were incubated over ice for 40 min. They were then spotted onto parafilm placed over ice and exposed to UV light for 1.5 min using a UV Stratalinker 2400 (Stratagene) set on time mode. Seven microliters of 4× NuPAGE LDS sample buffer (Invitrogen) were added to 20 μl of each reaction, and the samples were boiled for 5 min and electrophoresed in 4–12% Bis-Tris NuPAGE gels (Invitrogen). Gels were subjected to autoradiography. For the quantitative UV cross-linking assays, L2, 6, and 30 pmol of proteins were combined with 1 pmol of DNA.

**RESULTS**

**The Domain Structure of Tf1 IN**—The GP(Y/F) domain was reported to be a conserved module in the INs of the Metaviridae family of LTR retrotransposons and in the γ family of retroviruses (10, 11). The alignment of INs in Fig. 1A shows the broad conservation of the GP(Y/F) domain among these distantly related families. The amino acids of Tf1 IN that encompass the GP(Y/F) domain are aa 339–368 (10), and this is the definition of the Y/F that defines the GP(Y/F) domain.

The domain was probed with anti-IN rabbit antibody (1:10,000) (41). The secondary antibody was horseradish peroxidase-conjugated donkey anti-rabbit Ig, whole antibody (1:10,000; Amersham Biosciences). ECL Plus was used to detect the protein signals (Amersham Biosciences).
architecture typical of INs. A scaled drawing of Tf1 IN indicates that the position of its HHCC and DDE motifs are quite similar to those of M-MuLV IN (Fig. 1B). To map the position of its domains, Tf1 IN purified from bacteria was subjected to partial proteolysis with trypsin, and the N termini of the products were sequenced. The results confirmed that the IN does have a three-domain structure with an N-terminal domain (aa 1–110), a catalytic core (aa 111–354), and a C-terminal domain (aa 355–477) (Fig. 1B and supplemental Fig. S1). Interestingly, the cleavage between the central and C-terminal domains occurred in the middle of the GP(Y/F) domain. This indicated that the GP(Y/F) domain assembles into two stable segments split by protease-accessible residues.

Substitutions in the GP(Y/F) Domain Significantly Reduce Transposition Activity—To test whether the amino acids of the GP(Y/F) domain possessed an important function, we made substitutions in the GPF residues of the transposon and measured the resulting transposition activity in vivo (Fig. 2). The assay for transposition activity consisted of expressing neo-containing copies of Tf1 in S. pombe and measuring the resistance to G418 that results from integration (17). The transposition frequencies of elements with the substitutions G364A, P365A, F366A, and G364A/P365A/F366A ((364–366)AAA) were significantly reduced (Fig. 2A). When the proportion of the cells with resistance to G418 was quantified, we found the substitutions G364A, P365A, F366A, and G364A/P365A/F366A reduced transposition by 59-, 7.6-, 26-, and 72-fold, respectively (supplemental Table S3). In order to analyze which steps of transposition were disrupted by the substitutions, cDNA synthesis and levels of IN were analyzed. Immunoblots revealed that the substitutions significantly reduced the levels of IN expressed in S. pombe (Fig. 2B). These reduced levels of IN made it difficult to identify a specific function of the GP(Y/F) domain in in vivo assays. DNA blots of cells expressing Tf1-neo revealed that the substitutions caused no more than a 2-fold defect in cDNA production (supplemental Fig. S2A). In addition, the substitutions did not reduce levels of reverse transcriptase (supplemental Fig. S2B).

The Role of the GP(Y/F) Domain in Disintegration and Strand Transfer—To determine whether the amino acids in the GP(Y/F) domain contributed to the catalytic activity of IN and to compare its function to other domains, we purified a set of Tf1 INs that contained various truncations (Fig. 3A). Each of the recombinant Tf1 proteins was assayed for disintegration activity with a previously established substrate that mimics the U3 end of the LTR inserted in a target site (Fig. 4) (16). By itself, the catalytic core domain of
Tf1 IN lacked activity (Fig. 4B, lanes 9–12). As observed previously, IN lacking the CHD (aa 1–406) was considerably more active than the full-length IN (Fig. 4B, lanes 1–8) (16). Truncations revealed that the N-terminal domain was required for catalytic activity (Fig. 4C, lanes 4–8). Other deletions revealed that the C-terminal domain was also necessary for activity (Fig. 4C, lanes 9 and 10). This requirement of the C-terminal domain is consistent with studies of the M-MuLV IN that found a similar requirement for the C terminus using disintegration assays (20). However, the same study found that the N-terminal domain of M-MuLV IN was not required for activity. Thus, Tf1 IN is distinct from the IN of M-MuLV in that it requires both N-terminal and C-terminal domains for disintegration activity.

The C-terminal deletions removed the GP(Y/F) domain as well as downstream amino acids. To examine the role specifically of the GP(Y/F) domain, we substituted each of the GPF residues to alanine (G364A, P365A, and F366A). Unfortunately, full-length INs with the substitutions G364A and F366A were insoluble and could not be purified. Nevertheless, we were able to isolate full-length IN with the P365A substitution (Fig. 3B, bottom left).

The isolation of full-length IN with the P365A substitution provided the opportunity to ask whether a single amino acid substitution in the GP(Y/F) domain would affect the catalytic activity of IN. When assayed with the disintegration substrate, the P365A substitution caused a moderate reduction in activity (Fig. 5A). However, the disintegration assay monitors strand joining in an intramolecular reaction that has modest substrate specificity. The strand transfer assay with oligonucleotides that mimic the double-stranded end of an LTR is a stringent test of an IN's ability to complete strand joining of two molecules of substrate DNA. We therefore used the strand transfer assay to determine whether the P365A substitution reduced this integration activity. Relative to wild-type IN, the IN with the P365A substitution had significantly less strand transfer activity (Fig. 5B).

Quantification of the disintegration assays revealed that the P365A substitution caused a 3.6-fold reduction in product when measured at the protein concentration that gave the highest activity of wild-type IN, 0.3 μM (Fig. 5C, left). Surprisingly, the P365A substitution caused the IN concentration with maximum activity to increase 4-fold to 1.2 μM. At the optimal concentrations of both INs, the protein with the P365A substitution had 55% of the activity of wild-type IN. In contrast to this modest reduction in disintegration activity, P365A caused a 13.4-fold decrease in strand transfer activity when measured at an IN concentration of 0.15 μM, the optimal concentration of both INs.
The GP(Y/F) Residues Do Not Contribute to the DNA Binding Activity in the C Terminus of Tf1 IN—To determine how the GP(Y/F) domain contributed to strand transfer activity, we tested how substitutions in the GP(Y/F) residues altered properties associated with IN. The C-terminal domains of INs are known to bind DNA without any sequence specificity (1). To determine whether the GP(Y/F) domain contributed to DNA binding, we first mapped which sections of Tf1 IN interacted with DNA. Labeled oligonucleotides with sequence from the U5 end of the LTR were mixed with 1/9262 g of various proteins, and the mixtures were cross-linked with UV. DNA binding activity was observed with the full-length IN, CHD, and the fragment consisting of amino acids 335–406 (Fig. 6A). This fragment (aa 335–406) is referred to here as the GP(Y/F) fragment, because it included the entire GP(Y/F) domain. The proteins with the strongest DNA binding were CHD, the catalytic core, and the GP(Y/F) fragment. The CHD of the retrotransposon Maggy was used as a negative control, because it is similar to the CHD of HP1 and binds specifically to histone H3 methylated at lysine 9 (12). The Tf1 CHD exhibited...
The GP(Y/F) Domain of IN

conduct a series of experiments to determine whether the GP(Y/F) domain forms multimers and to compare its ability to multimerize to that of the other domains. In initial experiments to test the domains of Tf1 IN for the propensity to multimerize, we tested full-length IN for interactions with the individual portions of the protein. Using our set of recombinant proteins in precipitation experiments, we found that the N-terminal domain, the central core, and the 71-amino acid GP(Y/F) fragment (aa 335–406) all bound to the full-length IN (supplemental Fig. S4). Interestingly, removing 20 amino acids from the N terminus of the fragment with the GP(Y/F) domain resulted in a protein (GP(Y/F)ΔN) that did not bind IN. This suggests that the N-terminal half of the GP(Y/F) domain played an important role in binding IN.

In order to test whether the GP(Y/F) domain contributes to multimerization, we subjected the IN proteins to gel filtration on Superdex 200 using protein concentrations of 1 mg/ml and a buffer of 50 mM HEPES, pH 7.5, 0.5 M NaCl, and 1% (v/v) glycerol. Full-length wild-type IN eluted with an apparent mass of 119.5 kDa, indicating that it formed a stable dimer (Fig. 7A). In efforts to detect a tetrameric form, gel filtration was performed in 1 M NaCl with protein concentrations of 1.0 and 2.0 mg/ml. Under these conditions, HIV-1 IN is in equilibrium between dimers and tetramers (28). However, only dimer species of Tf1 IN were observed (data not shown).

We tested whether the GP(Y/F) domain was required for dimerization by subjecting the full-length IN with P365A to gel filtration. The altered protein eluted with an apparent mass of 103.8 kDa, indicating that the P365A substitution did not reduce the formation of dimers (Fig. 7A, bottom).

Since the N-terminal, catalytic core, and C-terminal domains of INs each form stable dimers when tested as individual fragments (1, 26, 27), such interactions had the potential to mask any contribution the GP(Y/F) domain may have made to the dimerization of the full-length IN. To determine whether this was possible with Tf1 IN, we performed gel filtration on the catalytic core (aa 110–354) and the IN lacking the C-terminal domain, ΔC (Fig. 7B). Because these experiments were run separately from those in Fig. 7A, a set of molecular weight standards and full-length IN were run to calibrate the column. Both the core and the ΔC had apparent weights indicative of stable dimers. As a result, any contribution made by the GP(Y/F) domain to multimerization could have been masked by the core and N-terminal domains.

To test directly whether the individual fragments of the C-terminal domain promoted multimerization, gel filtration with Superdex 75 was performed. At a concentration of 1.0 mg/ml, the CHD (aa 407–477) eluted as a monomer (Fig. 8A). Surprisingly, the profile produced by the GP(Y/F) fragment (aa 335–406) included three major peaks (Fig. 8B). The apparent size of these species indicated the presence of monomer, dimer, and trimer. These results were interesting because they indicated that the small GP(Y/F) domain itself was capable of forming multimers larger than dimers. To test whether the highly conserved GP residues of the GP(Y/F) domain contributed to this multimerization, GP(Y/F) fragments with single amino acid substitutions were analyzed. Both substitutions, G364A and P365A, disrupted all multimerization of the GP(Y/F) frag-
ment (Fig. 8, C and D). These data indicate that the GPF residues played an important role in promoting multimerization of the GP(Y/F) fragment.

Gel filtration of the GP(Y/F) fragment did not resolve multimers larger than trimers. To test for larger multimers, we subjected the GP(Y/F) fragment to the chemical cross-linker bis(sulfosuccinimidyl)suberate. Gel electrophoresis of the cross-linked samples indicated the protein at a concentration of $25\mu{M}$ formed an equilibrium of monomers, dimers, trimers, and tetramers (Fig. 9). Thus, this 71-amino acid fragment containing the GP(Y/F) domain was able to form multimers as large as tetramers.

**DISCUSSION**

Trypsin proteolysis revealed that Tf1 IN possessed the three domain architecture typical of retrotransposon and retrovirus INs. In addition to this general property of INs, the IN of Tf1 has specific features that are similar to the IN of M-MuLV.
These INs have HHCC and DDE motifs in positions that are closely matched between the two proteins. Of particular interest is that both Tf1 and M-MuLV INs have GP(Y/F) domains in very similar positions. These observations indicate that the activities of the GP(Y/F) domain in Tf1 IN may reflect the function of the GP(Y/F) domain of M-MuLV as well as other γ-retroviruses.

Our in vivo assays of transposon function revealed that substitutions of the Gly, Pro, and Phe residues in the GP(Y/F) domain caused IN to become unstable. Although the reduction in the levels of IN made it difficult to evaluate the function of these amino acids, it did suggest that the GP(Y/F) domain was an important structural feature needed for the protein to fold correctly. Regardless of their effect on integration, the substitutions in the GP(Y/F) domain had little impact on the stability of reverse transcriptase and the production of cDNA. This finding reflects previous observations that reverse transcriptase produces normal levels of cDNA even when Tf1 lacks IN expression (Fig. S2, INΔ6) (29).

For in vitro analysis, the disintegration assay is a sensitive method for identifying the minimum structure of an IN that is capable of performing strand breaking and joining. With this assay, it was shown that the central core domains of HIV-1 and Rous sarcoma virus INs possess all of the active site and substrate binding residues necessary for completing catalysis (18, 19). In the case of the M-MuLV and Tf1 INs, these enzymes have larger C-terminal domains, and these domains are required to catalyze disintegration; for Tf1, the N-terminal domain was also necessary. The requirement for these additional domains is not understood, but it is possible that they stabilize binding to the disintegration substrate.

The segment of the C terminus in Tf1 IN that was required for disintegration activity was amino acids 354–406. This portion of the C terminus contained the GP(Y/F) domain as well as much of the residues in the DNA binding GP(Y/F) fragment. Since the P365A substitution in the full-length IN caused only a modest reduction in disintegration, we propose the possibility that the GP(Y/F) residues themselves were not the component of the C-terminal domain that was essential for disintegration activity. Instead, its DNA binding activity may have provided the critical function removed by the C-terminal truncation of amino acids 354–406. This model is consistent with our finding that the DNA binding activity in the C-terminal domain functioned independently of the GP(Y/F) residues.

Although the P365A substitution did not cause a substantial reduction in disintegration activity, it did increase the concentration of IN required for maximal activity. It is not known what is responsible for the inhibition of activity caused by high concentrations of IN. Nevertheless, it is possible that P365A weakened subunit interactions necessary for disintegration. The higher concentrations of IN would be required to compensate for the defects in multimerization. However, other explanations are also possible.

In strand transfer assays, the P365A substitution caused a much greater defect in activity than was observed in disintegration reactions. This suggests that the GP(Y/F) domain provided a key function in strand transfer that was less important in disintegration assays. Functions of IN that contribute more significantly to strand transfer than disintegration include recognition of the terminal sequences of donor DNA as well as the ability to join two separate DNA substrates (30). It may be that Pro-365 played a disproportional role in one of these functions.

In order to distinguish between the possible roles of the GP(Y/F) domain, we studied the biochemistry of a 71-amino acid fragment containing the GP(Y/F) domain. In our UV cross-linking assays with oligonucleotides, the GP(Y/F) fragment did bind DNA with high efficiency. This finding that a region of the C-terminal domain bound DNA and that it was a nonspecific DNA binding activity paralleled the activities of other INs (31–36). However, our study of single amino acid substitutions indicated that the GP(Y/F) residues within the GP(Y/F) fragment did not contribute to this DNA binding. Also of interest was our finding that IN lacking the CHD (CHDΔ) bound DNA more efficiently than the full-length IN (Fig. 6A). This observation suggests that the CHD can interfere with DNA binding. Such a reduction in DNA binding could explain why CHDΔ has greater catalytic activity than the full-length IN. This is consistent with the possibility that the CHD performs a regulatory function during integration.

Another property of INs important for activity is multimerization. Sedimentation and kinetic experiments indicate that the INs of avian sarcoma virus and HIV-1 must multimerize to become active (21, 37). Complementation studies with two defective forms of HIV-1 IN revealed that subunits can multimerize to become active (38, 39). Recent studies indicate that the IN of HIV-1 is a tetramer in its synaptic complex and that multimerization of the C-terminal domain plays an important role in concerted integration (7, 40). Our results of gel filtration indicated that full-length Tf1 IN formed a stable dimer. Despite a significant effort, we could not identify conditions that allowed IN to form tetramers. Nevertheless, complementation analyses with defective versions of Tf1 IN revealed that multimeric complexes were highly active.4

4 H. Ebina and H. Levin, unpublished results.
The GP(Y/F) Domain of IN

In experiments to test whether the GP(Y/F) domain contributed to multimerization, we found the substitution P365A in full-length IN did not diminish dimerization. We also found that stable dimers were formed by IN lacking the C-terminal domain (aa 1–334) and with the core domain (aa 110–354) itself. Thus, we do not have direct evidence that the GP(Y/F) domain contributes to multimerization of the full-length protein. It is possible that the GP(Y/F) domain does not promote dimerization of IN. However, any contribution that the GP(Y/F) domain might make to dimerization could have been masked by self-association interactions in other regions of the protein.

Our direct examination of the GP(Y/F) fragment by gel filtration and chemical cross-linking revealed high levels of dimers, trimers, and tetramers. This ability to form trimers and tetramers was unique among the INs and IN fragments we studied. The multimerization of the GP(Y/F) fragment was disrupted by the single amino acid substitutions in Gly-364 and Pro-365, indicating that the GP(Y/F) domain played a significant role in multimerization. It was interesting that these same substitutions did not compromise the DNA binding activity of the GP(Y/F) fragment. We therefore concluded that multimerization of the GP(Y/F) fragment was not necessary for the DNA binding activity.

The lack of any high resolution structure of TFI IN and our inability to detect its tetramerization makes it difficult to speculate about the role of the GP(Y/F) domain in the multimerization of the full-length IN. It is reasonable to propose that the dimerization of the GP(Y/F) domain contributes to the multimerization of the full-length protein. It is tempting to speculate that like the IN of HIV-1, TFI IN may form higher multimers in its synaptic complex. If this is true, the GP(Y/F) domain could mediate the higher multimerization. This speculation is consistent with the increase in IN concentration necessary for peak activity that was caused by the P365A substitution. Whether the drop in strand transfer activity caused by P365A resulted from a defect in tetramerization is not known. It is possible that the substitution resulted in other structural perturbations. Nevertheless, the significant drop in strand transfer activity caused by P365A suggests that the GP(Y/F) domain plays an important function in integration.

Although little sequence conservation of the GP(Y/F) domain is observed in the IN of HIV-1, the C terminus of TFI, avian sarcoma virus, and HIV-1 INs all bind DNA. Thus, it is possible that the C terminus of TFI IN adopts the Src homology 3-like folds present in the INs of HIV-1 and avian sarcoma virus. Nevertheless, the conservation of the GP(Y/F) domain in the Metavirus family of retrotransposons and in the diverse family of γ-retroviruses indicates that its key function is broadly conserved.

Acknowledgments—We thank Dr. Nga Nguyen (Food and Drug Administration Center for Biologics and Evaluation) for assistance with N-terminal sequencing. We are grateful to Amnon Hizi for providing antibodies against reverse transcriptase.

REFERENCES

1. Craigie, R. (2002) in Mobile DNA II (Craig, N. L., Craigie, R., Gellert, M., and Lambowitz, A. M., eds) pp. 613–630, American Society for Microbiology Press, Washington, DC
2. Fujiiwara, T., and Mizuuchi, K. (1988) Cell 54, 497–504
3. Engelman, A., Mizuuchi, K., and Craigie, R. (1991) Cell 67, 1211–1221
4. Sinha, S., Pursley, M. H., and Grandgenett, D. P. (2002) J. Virol. 76, 3105–3113
5. Chiu, R., and Grandgenett, D. P. (2003) J. Virol. 77, 6482–6492
6. Sinha, S., and Grandgenett, D. P. (2005) J. Virol. 79, 8208–8216
7. Li, M., and Craigie, R. (2005) J. Biol. Chem. 280, 29334–29339
8. Chow, S. A., and Brown, P. O. (1994) J. Virol. 68, 3896–3907
9. Chow, S. A., Vincent, K. A., Ellison, V., and Brown, P. O. (1992) Science 255, 723–726
10. Malik, H. S., and Eickbush, T. H. (1999) J. Virol. 73, 5186–5190
11. Lutzke, R. A., Plasterk, R. H. A. (1998) Biochemistry 37, 7712–7718
12. Lin, J. H., and Levin, H. L. (1997) Genes Dev. 11, 270–285
13. Bushman, F. D., Engelman, A., Palmer, I., Wingfield, P., and Craigie, R. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 3428–3432
14. Koloskowy, J., Katz, R. A., Merkel, G., and Skalka, A. M. (1995) Virology 206, 448–456
15. Jonsson, C. B., Donzella, G. A., Gauan, E., Smith, C. M., and Roth, M. J. (1996) J. Virol. 70, 4585–4597
16. Jones, K. S., Coleman, J., Merkel, G. W., Laue, T. M., and Skalka, A. M. (1992) J. Biol. Chem. 267, 16037–16040
17. Jenkins, T. M., Engelman, A., Ghirlanda, R., and Craigie, R. (1996) J. Biol. Chem. 271, 7712–7718
18. Zhang, S. P., Xiao, I. M., Knudson, J. R., Lewis, M. S., and Han, M. K. (1997) Biochemistry 36, 173–180
19. Coleman, J., Eaton, S., Merkel, G., Skalka, A. M., and Laue, T. M. (1999) J. Biol. Chem. 274, 32842–32846
20. Villanueva, R. A., Jonsson, C. B., Jones, J., Georgiadis, M. M., and Roth, M. J. (2003) Virology 316, 146–160
21. Dyda, F., Hickey, A. M., Jenkin, T. M., Engelman, A., Craigie, R., and Davies, D. R. (1994) Science 266, 1981–1986
22. Andrake, M. D., and Skalka, A. M. (1995) J. Biol. Chem. 270, 29299–29306
23. Lutzke, R. A. P., and Plasterk, R. H. A. (1998) J. Virol. 72, 4841–4848
24. Atwood, A., Choi, J., and Levin, H. L. (1998) J. Virol. 72, 1324–1333
25. Vincent, K. A., Ellison, V., Chow, S. A., and Brown, P. O. (1993) J. Virol. 67, 425–437
26. Khan, E., Mack, J. P., Katz, R. A., Koloskowy, J., and Skalka, A. M. (1991) Nucleic Acids Res. 19, 851–860
27. Vink, C., Groenevanger, A., and Plasterk, R. H. A. (1993) Nucleic Acids Res. 21, 1419–1425
28. Engelman, A., Hickman, A. B., and Craigie, R. (1994) J. Virol. 68, 5911–5917
29. Woerner, A. M., and Marcussekura, C. J. (1993) J. Biol. Chem. 268, 35913–35917
30. Vincent, K. A., Ellison, V., Brown, P. O., and Craigie, R. (2005) J. Biol. Chem. 280, 29334–29339
31. Lutzke, R. A., Vink, C., and Plasterk, R. H. A. (1994) Nucleic Acids Res. 22, 4125–4131
32. Esposito, D., and Craigie, R. (1998) EMBO J. 17, 5832–5843
33. Engelman, A., Hickman, A. B., and Craigie, R. (1994) J. Biol. Chem. 269, 29279–29287
34. Woerner, A. M., and Marcussekura, C. J. (1993) Nucleic Acids Res. 21, 3507–3511
35. Lutzke, R. A., Vink, C., and Plasterk, R. H. A. (1994) Nucleic Acids Res. 22, 4125–4131
36. Esposito, D., and Craigie, R. (1998) EMBO J. 17, 5832–5843
37. Hickman, A. B., Palmer, I., Engelman, A., Craigie, R., and Wingfield, P. (1994) J. Biol. Chem. 269, 29279–29287
38. Engelman, A., Bushman, F. D., and Craigie, R. (1993) EMBO J. 12, 3269–3275
39. Vangent, D. C., Vink, C., Groenevanger, A., and Plasterk, R. H. A. (1993) EMBO J. 12, 3261–3267
40. Li, M., Mizuuchi, M., Burke, T. R., and Craigie, R. (2006) EMBO J. 25, 1295–1304
41. Levin, H., Weaver, D. C., and Boeke, J. D. (1993) EMBO J. 12, 4885–4895