The Integrase of the Long Terminal Repeat-Retrotransposon Tf1 Has a Chromodomain That Modulates Integrase Activities*

Amnon Hizi1 and Henry L. Levin2

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

Chromodomains in a variety of proteins mediate the formation of heterochromatin by interacting directly with histone H3, DNA, or RNA. A diverse family of long terminal repeat (LTR)-retrotransposons possesses chromodomains in their integrases (IN), suggesting that the chromodomains may control integration. The LTR-retrotransposon Tf1 of Schizosaccharomyces pombe is highly active and possesses a chromodomain in the COOH terminus of its IN. To test this chromodomain for a role in integration, recombinant INs with and without the chromodomain were assayed for activity in vitro reactions. The full-length IN had integration activity with oligonucleotide substrates that modeled both the insertion and a reverse reaction known as disintegration. The INs of retroviruses possess an additional activity termed 3′ processing that must remove 2–3 nucleotides from the 3′ ends of the viral cDNA before insertion can occur. These additional nucleotides are added during reverse transcription because of the position of the minus strand primer downstream of the LTR. The position of the primer for Tf1 suggests no nucleotides are added 3′ of the LTR. It was therefore surprising that Tf1 IN was capable of 3′ cleavage. The most unexpected result reported here was that the IN lacking the chromodomain had significantly higher activity and substantially reduced substrate specificity. These results reveal that both the activity and specificity of enzymes can be modulated by their chromodomains.

Post-translational modifications of histones, such as the acetylation or methylation of lysines, establish the accessibility of DNA to the transcription machinery (1–3). Conserved domains that recognize these specific modifications exist in a variety of factors that interact with the histones and propagate the structures of chromatin (4–6). A large family of proteins that regulate chromatin structure bind specifically to histone H3 with methylated lysines using a conserved module called a chromodomain (6, 7). Although all the chromodomains studied thus far function to regulate chromatin structure, sequence analyses revealed large families of long terminal repeat (LTR)-retrotransposons with chromodomains in the COOH termini of their INs (8, 9). This is intriguing because it suggests that the transposons acquired chromodomains to regulate integration.

LTR-retrotransposons are closely related to retroviruses. Both types of elements encode Gag, protease, reverse transcriptase, and IN proteins and both depend on IN to insert their cDNA into the genome of a host cell. The LTRs, positioned at the ends of the cDNA, are subdivided into an upstream segment, U3, a central region, R, and a downstream segment, U5. A large number of biochemical studies with purified INs have characterized the chemistry of insertion (10–12). The INs of retroviruses have two activities that are necessary for the integration of the viral cDNA. At first, 2 or 3 nucleotides are removed from the 3′ ends of the LTRs. This site-specific nucleolytic cleavage or “3′ processing reaction” is required to position the “CA” dinucleotides at the 3′ ends of the cDNA. The CA dinucleotides are conserved among all retroviruses and their position at the 3′ ends of the cDNA is required for the second step of retrovirus integration, the “strand transfer reaction.” During this second step, the 3′ hydroxyls of the terminal “A” act as nucleophiles in a pair of transesterification reactions that cleave the phosphodiester bonds in the target DNA and simultaneously make a covalent bond between the 3′ ends of the viral DNA and the 5′ ends of the target DNA.

In retroviruses, the minus strand primers initiate reverse transcription 2 or 3 nucleotides from the LTR and this results in additional nucleotides at the cDNA termini that IN removes during the processing reaction. In the case of most LTR-retrotransposons the minus strand primers initiate reverse transcription with no added nucleotides. As a result, these LTR-retrotransposons are thought to proceed to the strand transfer reaction without a processing step (13).

The INs of retroviruses have three protease-resistant domains and the functions of these domains have been studied extensively (12, 14, 15). The NH2-terminal domain contains a conserved motif similar to zinc fingers. Although this domain does bind zinc, little else is known about its function. The central domain is the catalytic core that contains highly conserved acidic residues, Asp, Asp, and Glu, that are juxtaposed on an RNase H-like fold. These three amino acids are conserved in all the INs of retroviruses, LTR-retrotransposons, and DNA transposons (16). The central domain of HIV-1 IN was shown to be sufficient for catalytic activity in in vitro assays for the reverse reaction of strand transfer known as disintegration (17). The COOH-terminal domain is less well conserved but has nonspecific binding activity to DNA. Although the structure of individual domains has been determined for several INs, no single structure yet contains all three domains (18).

* This work was supported in part by the Intramural Research Program of the National Institutes of Health, NICHD, and the National Institutes of Health Intramural AIDS Targeted Antiviral Program. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact. We dedicate this article to the memory of our colleague and good friend, Dr. Kiebarg Nam.

1 Recipient of an Oak Ridge Institute for Science and Education fellowship administered by Oak Ridge Laboratories. Permanent address: Dept. of Cell and Developmental Biology, Sackler School of Medicine, Tel Aviv University, 69978 Israel. Incumbent of The Gregorio and Dora Shapira Chair for the Research of Malignancies at Tel Aviv University. E-mail: ahizi@post.tau.ac.il.

2 To whom correspondence should be addressed: 18 Library Dr., Rm. 106, Bethesda, MD 20892. Tel.: 301-402-4281; Fax: 301-402-1323; E-mail: henry_levin@nih.gov.

3 The abbreviations used are: LTR, long terminal repeat; IN, integrase; CH3, the chromodomain-deleted IN; HIV-1, human immunodeficiency virus type 1; MeSO2, dimethyl sulfoxide; nt, nucleotide(s); RT, reverse transcriptase; Ov, overhang; MOPS, 4-morpholinepropanesulfonic acid.
**The Effect of the Chromodomain on the Integrase of Tf1**

**Tf1 of Schizosaccharomyces pombe** is one of the LTR-retrotransposons that contains a chromodomain at the COOH terminus of IN (8). Its high level of transposition activity allows Tf1 to be studied as a model for reverse transcription and integration. Expression of Tf1 from a high copy plasmid results in the formation of virus-like particles, the reverse transcription of the Tf1 mRNA, and the integration of the cDNA into intergenic sequences (19–21). A novel mechanism of self-priming initiated minus strand reverse transcription from the upstream LTR (22). After initiation, reverse transcription proceeds through the same intermediates typical of other LTR elements. The IN of Tf1 that inserts this cDNA into the genome of *S. pombe* contains putative zinc binding, catalytic core, and COOH-terminal domains that are necessary for integration (23). Mutation analysis indicated that its high level of activity is sufficient to predict the presence of processing activity. This indicates that the position of the minus strand primer relative to the LTR is not sufficient to predict the presence of processing activity. The most unexpected result was that CH− had significantly higher specific activity than IN. In addition, CH− was much less dependent on specific sequences at the termini of the donor DNA than was IN. These results reveal that the chromodomain in the Tf1 IN modulates both the level and specificity of integration.

**EXPERIMENTAL PROCEDURES**

**Plasmid Constructions and Cell Growth**—The construction of pHL2468, the plasmid for the expression of the full-length Tf1 IN has been previously described (19). Briefly, the IN-coding DNA segment of Tf1 was amplified by PCR. The resulting fragment was flanked with Sail sites that were introduced into the BamHI site of pET15B and sequenced. The NH2 terminus of the Tf1 IN was predicted based on the amino terminus of the IN from a closely related element, Ty3 (24). The NH2 terminus of the recombinant Tf1 IN started with the sequence -VHMSM (25) and the COOH terminus ended with the sequence -NNLNI. The resulting protein was 477 residues long and was fused at its amino terminus of the recombinant Tf1 IN with a 21-residue peptide containing the His6 tag and the thrombin cleavage site. The plasmid expressing CH−, pH2469, was generated by a similar method. The CH− coding DNA segment was amplified by PCR with the primers HLI167 and HL1168, using the *Pfu* Turbo thermostable DNA polymerase (Stratagene). The DNA generated was cleaved with NdeI and BamHI, cloned into the vector pET15B, and sequenced. The CH− protein had the same amino terminus as the full-length IN but ended with the sequence -RNHSE. Altogether CH− was 406 + 21 (tag) residues long. Both plasmids were introduced into the strain BL21. The bacteria were grown in LB broth supplemented with 100 μg/ml of carbenicillin. The culture temperature was 32 °C and the cells were grown until they reached an absorbance of about 0.6. Induction was performed with 1 mM isopropyl-1-thio-β-d-galactopyranoside for an additional period of 3 h at 32 °C. The cells were harvested, and the pellets were stored at −80 °C.

**Purification of the Recombinant Full-length IN and CH− Proteins**—All steps were performed at 4 °C, unless otherwise stated.

**Lysis of the Bacteria**—The bacterial pellets were brought up in 50 mM sodium phosphate buffer, pH 8.0, 0.1 M NaCl, 1 mM phenylmethylsulfonil fluoride, and 0.5 mg/ml lysozyme. After a thorough mixing, the solutions were kept for 15 min on ice followed by the addition of more NaCl to a final concentration of 0.5 M. Then, the solutions were sonicated on ice, followed by the addition of Triton X-100 to a final concentration of 0.1% (v/v), and gentle mixing. We have found that the presence of this detergent is critical to dissociate the recombinant INs from the bacterial DNA (data not shown). The insoluble material was removed by ultracentrifugation for 1 h in the Beckman SW28 rotor at 26,000 rpm and the clear yellowish supernatants were subjected to affinity purification.

**Affinity Purification on Sepharose-Co2**+ Columns—Sepharose-Co2**+** columns (BD Talon, BD Biosciences) with bed volumes of 5 ml were equilibrated with 50 mM sodium phosphate, pH 8.0, 0.5 M NaCl, 10% (v/v) glycerol, and 0.1% Triton X-100, and then loaded with the supernatant from the ultracentrifugation. The column was then washed twice, once with 50 mM sodium phosphate, pH 7.5, 0.5 M NaCl, 10% glycerol, 0.1% Triton X-100, and the second with the same buffer, supplemented with 20 mM imidazole. Elution was performed with the same buffer but with a final concentration of 50 mM imidazole. The fractions were all analyzed by SDS-PAGE, and the peak IN protein fractions were pooled and dialyzed against 50 mM imidazole. The fractions were then washed with about 100 ml of the same buffer used for dialysis, and then loaded with the dialyzed protein solution. The same buffer was also used for extensive washing (about 500 ml) of the column after loading was completed. Finally, the proteins were eluted off the affinity resin with a similar buffer, supplemented with NaCl at a final concentration of 0.4 M. The recombinant INs emerged off the columns at sharp peaks with protein concentrations as high as 1.5 mg/ml. The proteins were stored at −80 °C in small aliquots. Each aliquot was thawed only once for conducting the enzymatic assays.

**Oligonucleotides**—All oligonucleotides were gel purified. Their sequences and designations are presented in the supplemental material.

**Substrates Used in IN Assays**—The appropriate DNA oligonucleotides were 5′ end-labeled. Aliquots containing 0.4 μg of each oligo were 32P end-labeled using 20 units of T4 polynucleotide kinase (New England Biolabs), and 100 μCi of γ-32P[ATP in a final volume of 20 μl using the buffer supplied. After an incubation of 1 h at 37 °C, 1 μl of 0.25 mM 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 three times the equimolar ratio of the appropriate non-labeled DNA. After 5 min at 95 °C, the mixtures were allowed to cool slowly 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.

**Assays for IN Activities**—Unless otherwise stated, assays of IN and CH− were performed in final volumes of 20 μl containing 25 mM MOPS-NaOH, pH 7.2, 4% (w/v) polyethylene glycol 8000, 10% (v/v) glycerol, 5 mM dithiothreitol, 6.25% (v/v) of freshly prepared dimethyl sulfoxide, and 1 mM MnCl2. Each reaction mixture contained between 0.2 and 1 μg of the enzymes and 1 pmol of the labeled DNA substrates. Incubations were performed for 1 h at 37 °C, and reactions were stopped by adding 10 μl of formamide loading buffer (95% formamide, 20 mM EDTA, 0.5 mg/ml bromphenol blue, and 0.5 mg/ml xylene cyanole FF) to each sample. The samples were then treated for 3 min at 95 °C, quickly cooled on ice and loaded on an 8 M urea, 13.5% (w/v) polyacrylamide sequencing gel in Tris borate/EDTA buffer. After electrophore-
The Effect of the Chromodomain on the Integrase of Tf1

The wet gels were subjected to phosphorimaging, using an Amer sham Biosciences Storm 840 instrument.

Purified recombinant HIV-1 IN (a generous gift of R. Craigie) was assayed with the specified substrates under the optimal conditions of this enzyme that differ from those suitable to the Tf1 enzymes. Reactions were performed each in 20 μl of 25 mM MOPS-NaOH, pH 7.2, 0.1 mg/ml bovine serum albumin, 10% glycerol, 10 mM β-mercaptoethanol, 0.1 M NaCl, and 7.5 mM MnCl2 using 0.25 μg of the purified IN. All other steps, including preparation of the substrates, incubation times, and analyses of the reaction products were performed as described for the Tf1 INs.

RESULTS

Recombinant IN and CH⁺ of Tf1 Have Significant Levels of Disintegration Activity—To investigate the function of the chromodomain in Tf1 IN we studied the integration activities of the full-length IN and of a truncated IN that lacked the chromodomain (CH⁺). Both proteins were expressed with His₆ tags in bacteria and were purified to a high degree under “Experimental Procedures.” One μg of each purified protein and a mixture of size markers (M) containing 1 μg of protein in each band were loaded onto a 10% polyacrylamide-urea gel. Following electrophoresis the proteins were stained with Coomassie Brilliant Blue R-250.

Under in vitro conditions INs can catalyze the reverse of DNA strand transfer in a reaction designated disintegration (10). This activity can be assayed using oligonucleotides that mimic a single ended insertion (25–27). The disintegration activity of Tf1 IN was studied first because the chromodomain in HIV-1 IN greatly inhibited the activities of both proteins and a strong preference for MnCl₂ over MgCl₂ was observed. The highest disintegration activity was obtained with 25 mM MOPS-NaOH, pH 7.2, 4% polyethylene glycol 8000, 10% glycerol, 5 mM dithiothreitol, 6.25% Me₂SO, and 1 mM MnCl₂. Other additives, often used to enhance IN activities, such as, bovine serum albumin, spermidine, Triton X-100, or Me₂SO at final concentrations above 6.25% were all found to inhibit the Tf1 INs (data not shown).

Having established the assay conditions for the Tf1 IN and the CH⁺ enzymes, dose-dependent reactions were performed with increasing amounts of each protein. The quantitative data presented in Fig. 4 show that for the U3 substrate the activity of IN increased with protein content up to 275 ng, whereas no maximum was observed with the U5 substrate. The data with CH⁺ revealed the surprising result that the CH⁺ protein was substantially more active than the full-length IN. Although this was true for both the U3 and U5 substrates, it was with the U3 substrate that CH⁺ exhibited the greatest increase in activity over that of IN.

Recombinant IN and CH⁺ Possess Strand Transfer Activities—Strand transfer assays are more relevant measure of IN activities because they represent the forward reaction of integration. The strand transfer activities of IN and CH⁺ were assayed to determine whether these proteins were capable of mediating integration. The substrates for these assays were double-stranded oligonucleotides designated to mimic either the U3 or U5 ends of Tf1. Initially, we used double-stranded oligonucleo-
tides with blunt ends that terminated with the CA dinucleotide (Fig. 5A). The oligonucleotide with the 3’/H11032 CA was labeled at its 5’/H11032 end with 32P. Strand transfer results in the insertion of the oligonucleotides into other identical molecules and as a result oligonucleotides are created that are both bigger and smaller than their original size (Fig. 5A). We found that the conditions established for the disintegration reaction are also optimal for the strand transfer activity of the Tf1 enzymes (data not shown).

Fig. 5B shows that relative to protein dose, both the IN and CH− generated oligonucleotide products longer than the original U3 and U5 substrates. The lengths of these products varied because the integration into the acceptor DNA was relatively random. However, it was apparent for both IN and CH− that with the U5 substrate there were preferred positions. Interestingly, a significant fraction of the extended products were longer than twice the size of the 24-nt substrate and products as long as 70-mers were observed. This indicates that both versions of Tf1 INs were capable of three or more integration events in which products of the first round of integration served as substrates for further integration events. To our knowledge, such multiple round integrations have not been reported for other INs.

We tested whether the chromodomain played a role in strand transfer by comparing the activity of IN to CH−. To evaluate the levels of strand transfer activity we calculated the amount of the oligonucleotide species longer than 24 nt as a percentage of the total amount of the labeled oligonucleotides. Fig. 5C shows that CH− was substantially more active than IN with both the U3 and U5 ends. The elevated activity of CH− correlated well with the results of the disintegration assays. However, the difference between the strand transfer activity of CH− and IN was more pronounced than observed for disintegration. The maximal differences in activity between CH− and IN were approximately 7- and 4-fold when observed with the U3 and U5 end-derived substrates, respectively. These results indicate that the presence of the chromodo-
main inhibits both the disintegration and strand transfer activities of Tf1 IN.

**CH**⁻ and IN Possess 3’ End Processing Activity—It is currently thought that the role of the 3’ end processing activity of INs is solely to remove the 2 or 3 nucleotides 3’ of the CA that result from the position of the minus strand primer 2 or 3 nucleotides downstream of the first LTR. The primer for the minus strand of Tf1 is positioned adjacent to the LTR so it is predicted that no processing activity is required (28, 29). However, sequences of Tf1 cDNA extracted from particles revealed the surprising result that 85% of the 3’ ends had one or more untemplated nucleotides.⁴ Because these nucleotides are added to the 3’ ends after the CA dinucleotide, they are expected to block integration. We therefore speculated that Tf1 IN would have a processing activity that would remove these nucleotides to expose the CA dinucleotides and allow integration.

We tested IN and CH⁻ for the 3’ processing activity using the assay already described for the strand transfer activity. In this case, the 5’ end-labeled substrates contained nucleotides positioned 3’ of the critical CA (Fig. 6A). Processing activity would cause a reduction in length of the labeled substrate after the extra sequences are removed from the 3’ end.

**Processing the U5 End of the LTR—**The IN and CH⁻ enzymes were tested for processing activity using a set of substrates that mimicked the U5 end of the LTR with sequences that extended past the conserved CA. Each of the oligonucleotide pairs were designed such that processing activity that removes nucleotides 3’ of the CA would produce the same 24-nt product. The 3’ end extensions were either in the form of a single-stranded overhang (OV) beyond the CA or they were base paired in the form of a blunt end (BL). The sequences of the extensions were chosen based on their presence at the 3’ ends of the DNAs that were isolated in vivo from particles. For example, the most prominent sequences observed after the U3 and U5 ends were “TT” and “AT,” respectively.⁴ In addition, longer stretches of “T”s were observed.

As indicated by the 24-nt product, both IN and CH⁻ exhibited processing activity that specifically removed the nucleotides 3’ of the CA (Fig. 6B). The processing activity of IN produced the 24-nt sequence as the dominant product from the substrates with “tt”, “at”, “ttt,” and “ttttt” as single-stranded extensions. In the case of a substrate with a BL end (“at”), IN generated the 24-nt species along with several other smaller products. These additional products are seen to a lesser extent with the other substrates and are likely because of nonspecific nuclease activity of IN. This type of nonspecific activity is observed with other INs (30). It is reported that a variety of INs possess an intrinsic DNA hydrolysis or alcoholysis activity that is responsible for nonspecific nucleolytic activity similar to what we observed (31).

For retrovirus INs, the product of the processing activity is the only substrate suitable for strand transfer. Therefore, reactions with efficient processing activity also exhibit strand transfer. The processing activity of Tf1 IN with either the “at” or “tt” extensions was high and strand

---

⁴ Atwood-Moore, A., Ejebe, K., and Levin, H. L. (2005) J. Virol., in press.

**FIGURE 4.** The disintegration activities of IN and CH⁻ in assays with increasing amounts of protein. The reactions were conducted with either the U3 or U5 substrates and increasing amounts of Tf1 IN or CH⁻. The proteins were diluted in the buffer used to store the proteins (50 mM HEPES-NaOH, 10% glycerol, 0.05% MgSO₄, 2 mM diithiothreitol, and 0.4 M NaCl, final pH 7.2) and equal volumes of the protein samples were added to each reaction. After the reaction, products were subjected to electrophoresis and the gels were scanned with a phosphorimagery.

**FIGURE 5.** The strand transfer activities of IN and CH⁻. A, a schematic diagram of the strand transfer assay of the Tf1 IN and CH⁻ enzymes with pre-processed substrates. B, the electrophoretic analyses of the products of the strand transfer reactions. For the U3 end, oligonucleotide HL1034 was 5’ end-labeled and annealed to oligonucleotide HL1033. For the U5 end, oligonucleotide HL1039 was 5’ end-labeled and annealed to oligonucleotide HL1040. Decreasing amounts of each enzyme were incubated with either substrate. The products were resolved electrophoretically. The lanes marked 0 contained a reaction incubated with no enzyme. The following are the lane numbers and the amounts of protein, expressed in ng, used for each reaction: for IN: 1, 1.1; 2, 0.55; 3, 0.28; 4, 0.14; 5, 0.07; 6, 0.035; for CH⁻ protein: 7, 0.4; 8, 0.2; 9, 0.1; 10, 0.05; 11, 0.03. The protein dilutions were done as described in the legend to Fig. 4. The size markers included a disintegration reaction performed with U5 substrate and Tf1 IN, as described in the legend to Fig. 2. Also included as markers was the dX174 DNA cleaved with HinfI and 5’ end-labeled. C, quantitative analyses of strand transfer activity. The radioactivity in the gel shown in B was scanned with a phosphorimagery. The radioactivity in all the bands larger than 24 nt was calculated as the percentage of the total radioactivity. The activities are expressed as the percent of conversion.
transferred was detected (Fig. 6B). The processing activity with the 3′- and 5′-nucleotide extensions was less efficient and no strand transfer was observed. Thus, the IN of Tfl was similar to INs of retroviruses in that processing was required for strand transfer.

The amounts of processed products accumulated by CH⁻ were similar to that of IN (Fig. 6B). However, CH⁻ generated substantially more products of strand transfer than IN. The levels of these strand transfer products greater than 24 nt were quantified. The activity of CH⁻ with all substrates was significantly higher than that of IN. For example, the activities with the tt-Ov was 57% and with at-Ov was 38% relative to that produced with the pre-processed substrate. The comparable figures for IN were 11 and 15%, respectively. Because the increase in strand transfer requires enhanced 3′ end processing, these data indicate that the 3′ end processing activity of CH⁻ is higher than that of IN.

The efficient processing of the U5 substrates suggested that the Tfl proteins specifically recognized and removed the extensions beyond the CA dinucleotide. To test whether other sequences in the substrate were recognized, IN and CH⁻ were assayed for activity using a substrate with the U5 sequence of HIV-1 (Fig. 6B, far right). The HIV-1 substrate was 21 nt long and had the physiological GT extension 3′ of the CA dinucleotide. Processing activity would reduce the 21-nt species to 19 nt. Compared with the processing activity of HIV-1 IN with this substrate, the Tfl proteins exhibited very little activity. This indicates that in addition to the CA dinucleotide, the Tfl proteins recognize sequence in the U5 substrate specific for Tfl. It was also seen that HIV-1 IN did not utilize the Tfl pre-processed substrates (data not shown).

Processing the U3 End of the LTR—Retroviral INs have processing activities that remove extensions 3′ of the CA from both the U5 and U3 ends of the cDNA. Because the Tfl proteins were found to have processing activity on the U5 end, we tested whether they could also process extensions from the U3 end of Tfl. The results show that both CH⁻ and IN can remove nucleotides 3′ of the CA in the U3 sequence (Fig. 7). As seen with the U5 substrates, the shorter extensions on the U3 substrates were processed more efficiently than those with longer extensions. The cleavages of the Ov versions were somewhat more efficient than the Bl counterparts.

The processing activity of CH⁻ with 3′-, 5′-, and 6-nucleotide extensions was significantly greater than that of IN. However, comparing the activities of the proteins with the shorter extensions was complex because the strong strand transfer activity of CH⁻ reduced the accumulation of the processed species. To approximate the processing activities of IN and CH⁻, the levels of strand transfer were quantified. As found above for the U5 substrates, CH⁻ exhibited substantially higher levels of activity. For example, the activities with tt-Ov was 88%, with tt-BI was 32%, and with ttt-Ov was 19% of that produced with the pre-processed substrate. The comparable figures for IN were 11 and 4%, and not detectable, respectively.

Strand Transfer Activities Require Pre-processing of the Extensions 3′ to CA—The processing activities of IN and CH⁻ were estimated by measuring the products of the subsequent strand transfer. This was done because the products of processing can be readily converted by strand transfer activity into larger oligonucleotides. However, the accumulated products of strand transfer would be an overestimate of processing if the Tfl proteins could mediate strand transfer with substrates that retained nucleotide extensions 3′ to the CA. To test this possibility, strand transfer reactions were conducted with substrates that possessed a 2′,3′-dideoxynucleotide at their 3′ end. Unless removed by processing, this substrate would lack the necessary 3′ hydroxyl and as a result it would block strand transfer.

The experiment described in Fig. 8 tested the U3 substrate of Tfl that had 3′ ends with the sequence "ACA-ctt(dd)c." This substrate was compared with the control substrate that ended with a "c" instead of (dd)c. Two variants of each DNA, one with a single-stranded Ov, and the other with the Bl configuration were tested. The CH⁻ protein exhibited high levels of strand transfer with the blocked substrate ACA(t)c-ccc-Ov, which were not significantly different from that with the unblocked DNA, ACA(t)c-Ov. This indicated that the strand transfer occurred only after an efficient removal of the blocking sequence. The strand transfer activity with the Ov substrates was substantially greater than seen with the Bl versions. The full-length IN did not show significant activity with either the modified or unmodified substrates (Fig. 8).

The Strand Transfer Activity of Tfl IN but Not CH⁻ Specifically Requires That "ACA" Be Present at the 3′ End of the Substrate—The strand transfer activities of retrovirus and retrotransposon INs require that the highly conserved CA dinucleotide be present at the 3′ end of the
The Effect of the Chromodomain on the Integrase of Tf1

FIGURE 7. The 3’ processing and strand transfer activities of Tf1 IN and CH– with U3 substrates. The reactions were conducted with equal amounts of the specified substrates. The labeling of the substrates above each lane is as described in the legend to Fig. 6. The reactions performed with the specified enzymes are marked with + and those without enzyme are marked with –.

FIGURE 8. The 3’ end processing and strand transfer activities of Tf1 IN and CH– with substrates that contained blocked 3’ ends. Substrates ending with a 3’-terminal dideoxy C are indicated by ddc. The experiment was performed and the annotations are as described in the legend to Fig. 7.

donor substrate (30). We tested whether the IN of Tf1 had this same requirement for CA and whether the chromodomain contributed to this specificity. IN and CH– were tested for whether each of the last three nucleotides of the LTRs, ACA was necessary for strand transfer (Fig. 9). The substrates were based on the U3 double-stranded donor used in Figs. 7 and 8 that lacked extensions 3’ of the ACA. Thus, no 3’ processing was required. Single, double, and triple substitutions were made in the ACA to test the dependence on this triplet. All the modifications introduced in the sequence led to a dramatic reduction of the strand transfer activity of IN. In sharp contrast to IN, CH– was capable of strand transfer with most of the substituted substrates. Except for the substrate in which all three terminal nucleotides were flipped (tgg), CH– had greater than 10% of the activity observed with the unsubstituted substrate. The most striking difference between CH– and IN was with the substrate with a −1 transition (ATA). CH– retained 50% of the activity observed with the ACA terminus, whereas the full-length IN had just 1% activity. This large difference in the specificity of the two related proteins shows that IN was far more stringent than the CH– counterpart in selecting the DNA donor with the correct 3’ end.

DISCUSSION

In this study the IN of the LTR-retrotransposon Tf1 was produced as a recombinant, soluble and enzymatically active protein. Tf1 IN exhibited in vitro all the enzymatic activities expected for an IN from either a retrovirus or an LTR-retrotransposon. The presence of a chromodomain in the COOH terminus of Tf1 IN provided a unique opportunity to study the impact of such a domain on IN activity. Our results demonstrate that the chromodomain had a surprising influence on the specific activity of the IN as well as on its recognition of donor DNA.

Although the disintegration activity may not have direct biological relevance, it is particularly useful for biochemical characterization of INs. It provides a highly sensitive assay for the cleavage-ligation activity of IN. The disintegration assay developed here for Tf1 IN was used to optimize the reaction conditions. The increased activities of Tf1 IN with low concentrations of NaCl (Fig. 3A) were similar to what was observed with retroviral INs (26, 33). Because magnesium is thought to be the important divalent metal in vivo, the preference Tf1 IN had for manganese might be the result of structural defects in the recombinant protein. However, the preference for manganese could be a common property of INs as the INs of HIV-1 and murine leukemia virus also exhibit strong preferences for manganese over magnesium in vitro (26, 33). The ability of HIV-IN to react efficiently with the Tf1 substrate (Fig. 2) suggested that the disintegration activity of HIV-1 IN had little sequence specificity. This is consistent with data from several studies indicating that beyond the requirement for the CA, the disintegration activity of retroviral INs is not sequence specific (25, 34–36).

The full-length version of Tf1 IN was found to have substantial levels of strand transfer activity with both the U3 and U5 substrates. The genomes of eukaryotes accumulate large numbers of inactive retrotransposons. The strand transfer assay is a stringent test of IN function, and the activity of Tf1 IN detected here indicates that this protein was fully active. Strand transfer assays with substrates containing nucleotides 3’ to the CA revealed that the Tf1 IN also had 3’ processing activity (Figs. 6 and 7). This was observed with U3 and U5 substrates and with single and double-stranded extensions. As is the case with the 3’ end processing activity of HIV-1 IN, extensions longer than 2 nucleotides were more readily removed when they were single-stranded than when double-stranded (37). In one respect, the processing activity was quite unexpected. It was thought that the 3’ end processing activity of retroviruses and LTR-retrotransposons is required only in the cases where the priming of minus strand reverse transcription adds nucleotides beyond the conserved CA. For Tf1, the 3’ end of the minus strand primer is immediately adjacent to the U5 sequence of the upstream LTR (28, 29). As a result, no extensions and no processing were thought to occur. However, recent analyses of Tf1 cDNA extracted from particles revealed that 85% of the U3 and U5 ends terminated with nontemplated nucleotides. The processing activity of IN observed here could in theory remove these 3’ extensions and allow the bulk of the cDNA to participate in integration. This model raises the question why would nontemplated nucleotides be added just so they could be removed by IN. One possibility is that the nontemplated nucleotides could prevent aberrant annealing of the 3’ end to sequences in the opposite LTR and its subsequent extension. As a result, the nontemplated nucleotides would protect the completed cDNA until IN initiates 3’ processing. Another possibility is that the nontemplated nucleotides could protect the conserved CA from attack by nonspecific 3’ exonucleases.
The Effect of the Chromodomain on the Integrase of Tf1

The LTR-retrotransposon Ty3 of Saccharomyces cerevisiae does contain an IN with 3’ processing activity and this was predicted based on the position of its minus strand primer (38). Nevertheless, most other LTR-retrotransposons such as Ty1 have minus strand primers immediately adjacent to the U5 sequence of their LTR and are not thought to have processing (39). The finding that Tf1 IN has processing activity brings into question the predicted lack of processing activity in the INs of the other LTR-retrotransposons. Approximately 25% of Ty1 cDNA terminates with nontemplated nucleotides (40). The cDNA with the nontemplated additions could participate in integration only if the IN has processing activity. However, analyses of Ty1 IN failed to detect processing activity (41, 42).

The removal of nontemplated nucleotides by IN may also occur with retroviral cDNA. In vitro experiments show that HIV-1 reverse transcriptase possesses terminal transferase activity that adds nontemplated nucleotides to the 3’ end of the nascent DNA (43–45). A time course of reverse transcription during HIV-1 infection identified a prominent species of cDNA that had one nucleotide extra on the 3’ end of the plus strand (46). This led the authors to propose that the extra nucleotide was the result of nontemplated addition and that the 3’ processing activity of IN removed it.

Perhaps the most striking results presented in this report indicate a role for the chromodomain in modulating integration. Although removing 15% of IN might be expected to reduce activity substantially, the deletion of the chromodomain resulted in an intriguing increase in IN activity (47). This led the authors to propose that the extra nucleotide could participate in integration only if the IN has processing activity. However, analyses of Ty1 IN failed to detect processing activity (41, 42).

The poor solubility of the recombinant INs studied thus far has been a major obstacle in producing high resolution structures. High concentrations of salt or the introduction of amino acid substitutions have been necessary to solubilize individual domains of INs. A number of structures have been determined for each of the three domains of the INs from retroviruses HIV-1, simian immunodeficiency virus, avian sarcoma virus, and Roux sarcoma virus (18). However, the lack of a structure for an intact IN makes it difficult to model the spatial relationships between the three domains. During the purification of Tf1 IN it became clear that this IN was substantially more soluble than other integrases previously studied. It is possible that this solubility will lead to the first crystal structure of a full-length IN. This would serve as a significant leap in understanding the structures of INs in general and more specifically, of HIV-1 IN. A structure of an intact IN would substantially facilitate the design and development of new drugs against HIV IN and, thus, contribute to the ongoing fight against AIDS.

Acknowledgments—We express our gratitude to Dr. R. Craigie for helpful suggestions and fruitful discussions, providing purified HIV-1 IN and the substrates to assay it. We are very grateful to Drs. S. Chow and D. Grandgenett for valuable suggestions and J. Myung for help preparing the figures.

REFERENCES
1. Iizuka, M., and Smith, M. M. (2003) Curr. Opin. Genet. Dev. 13, 154–160
2. Kurdistani, S. K., and Grunstein, M. (2003) Nat. Rev. Mol. Cell. Biol. 4, 276–284
The Effect of the Chromodomain on the Integrase of Tf1

3. Fischle, W., Wang, Y. M., and Allis, C. D. (2003) Curr. Opin. Cell Biol. 15, 172–183
4. Hampsey, M., and Reinberg, D. (2003) Cell 113, 429–432
5. Robyr, D., Kurdistani, S. K., and Grunstein, M. (2004) in Methods Enzymol. 376, 289–304
6. Hampsey, M., and Reinberg, D. (2003) Cell 113, 429–432
7. Brehm, A., Tutfeland, K. R., Aasland, R., and Becker, P. B. (2004) Bioessays 26, 133–140
8. Malik, H. S., and Eickbush, T. H. (1999) J. Virol. 73, 5186–5190
9. Gorinsek, B., Gubensek, F., and Kordis, D. (2004) Mol. Biol. Evol. 21, 836–843
10. Craigie, R. (2001) J. Biol. Chem. 276, 23213–23216
11. Telesnitsky, A., and Goff, S. P. (1997) in Reverse Transcription and the Generation of Retroviral DNA, Cold Spring Harbor Press, Plainview, NY
12. Chiu, T. K., and Davies, D. R. (2004) Curr. Top. Med. Chem. 4, 965–979
13. Atwood, A., Lin, J. H., and Levin, H. L. (1996) Mol. Cell. Biol. 16, 338–346
14. Singleton, T. L., and Levin, H. L. (2002) Eukaryotic Cell 1, 44–55
15. Teysset, L., Dang, V. D., Kim, M. K., and Levin, H. L. (2003) J. Virol. 77, 5451–5463
16. Levin, H. L. (1995) Cell 88, 33–34
17. Bannister, A. J., and Kouzarides, T. (2004) in Methods Enzymol. 376, 269–288
18. Chow, S. A., Vincent, K. A., Ellison, V., and Brown, P. O. (1992) Science 255, 723–726
19. Vincent, K. A., Ellison, V., Chow, S. A., and Brown, P. O. (1993) J. Virol. 67, 425–437
20. Boeke, J. D. (1988) in RNA Genetics (Domingo, E., Holland, J. J., and Ahlquist, P., eds) Vol. II, pp. 59–103, CRC Press, Boca Raton, FL