Title
In vitro targeting of strand transfer by the Ty3 retroelement integrase.

Permalink
https://escholarship.org/uc/item/1j24f4pd

Journal
The Journal of biological chemistry, 287(22)

ISSN
0021-9258

Authors
Qi, Xiaojie
Sandmeyer, Suzanne

Publication Date
2012-05-01

DOI
10.1074/jbc.m111.326025

License
https://creativecommons.org/licenses/by/4.0/ 4.0

Peer reviewed
In Vitro Targeting of Strand Transfer by the Ty3 Retroelement Integrase*

Received for publication, November 30, 2011, and in revised form, March 13, 2012. Published, JBC Papers in Press, April 4, 2012, DOI 10.1074/jbc.M111.326025

Xiaojie Qi† and Suzanne Sandmeyer‡§

From the †Department of Biological Chemistry and ‡Institute for Genomics and Bioinformatics, University of California-Irvine, Irvine, California 92697

**Background:** Retroviruses and retrotransposons integrate with poorly understood preferences.

**Results:** Depending on in vitro reaction conditions, Ty3 retroelement integrase was targeted by a transcription factor or by Ty3 inverted-repeat terminal sequence.

**Conclusion:** Retroelement targeting can be reproduced in a defined system.

**Significance:** Biochemical models will enable improved understanding of retrovirus integration preferences and gene therapy vector safety.

The **Saccharomyces cerevisiae** long terminal repeat retrotransposon Ty3 integrates within one or two nucleotides of the transcription initiation sites of genes transcribed by RNA polymerase III. In this study the minimal components required to reconstitute position-specific strand transfer by Ty3 integrase are defined. Ty3 integrase targeted by a synthetic fusion of RNA polymerase III transcription factor IIB subunits, Brf1 and TBP, mediated position-specific strand transfer of duplex oligonucleotides representing the ends of the Ty3 cDNA. These results further delimit the TFIIB domains targeted by the Ty3 element and show that IN is the Ty3 component sufficient in vitro to target integration. These results underscore the commonality of protein interactions that mediate transcription and retrotransposon targeting. Surprisingly, in the presence of MnCl₂, strand transfer was TFIIB-independent and targeted sequences resembling the Ty3 terminal inverted repeat.

Retroviruses display preferential patterns of integration in eukaryotic genomes, reflecting influences of host transcription factors and effects of chromatin components and DNA modification, sequence, and structure on activity of the preintegration complex of integrase (IN)² and cDNA known as the intasome (1–3). Understanding these integration biases, particularly in the case of retroviruses, is complicated both by the complexity of the target and the animal genomes themselves. For example, although it is known that lens epithelium-derived growth factor (LEDGF) is required for efficient integration of HIV-1 (4–5) and that the interaction between IN and LEDGF maps to the C-terminal end of the catalytic core domain (6), mechanistic details of how LEDGF tethers the intasome to the target DNA remain elusive.

The relatively subtle integration preferences of retroviruses contrast with the striking preferences of some retrotransposons in lower eukaryotes and plants (7–9). For example, in **Saccharomyces cerevisiae**, the copia-like LTR retrotransposon Ty5 is targeted to heterochromatic DNA by interactions between the IN C-terminal domain and the Sir4 silencing protein (10–11) and copia-like Ty1 and gypsy-like Ty3 LTR retrotransposons target the 5′-flanking regions of Pol III-transcribed genes (12–13). In Dictyostelium discoideum, the non-LTR retrotransposon TRE5 targets 5′-flanking regions of tRNA genes and the TRE5 ORF1 protein interacts with components of the Pol III transcription factor TFIIB (14). The Schizosaccharomyces pombe gypsy-like Tf1 interacts with a subset of transcription factors to target RNA Pol II promoters (15).

Ty3 is distinguished by the precision of its integration within a few bases of Pol III TSS (13). However, despite this unusual insertion specificity, Ty3 has substantial structural and functional similarity to retroviruses (16). For example, cells expressing Ty3 accumulate VLPs containing processed Ty3 proteins and cDNA and the Ty3 IN has a conserved core domain that contains residues conserved among retroviral integrases, including the D, DX₁₅E catalytic motif of polynucleotide esterases. Although the amino- and carboxyl-terminal domains of IN proteins are generally less-well conserved, they contain a zinc finger and a GPY/F motif, respectively. These motifs are also found in the Ty3 IN protein (17). Similar to retroviral cDNA, the Ty3 cDNA has LTRs and terminates with two “extra” bp at each end, which are endonucleolytically removed from the 3′-ends by Ty3 IN prior to strand transfer (18). Based on the retrovirus model, the resulting 3′-hydroxyls mediate S_n2 nucleophilic attacks at staggered positions in the duplex chromosomal DNA. These positions are offset by 5 nts so that concerted strand transfer generates the characteristic 5-bp direct repeats flanking the ends of Ty3 insertions. These similarities to retroviruses coupled with precise targeting make Ty3 an attractive model for probing the mechanisms by which targeting proteins might interact with the retroelement intasome. However,
In Vitro Targeting of Strand Transfer by a Retroelement Integrase

a biochemically defined in vitro system that recapitulates the natural specificity of any retroelement including Ty3 has been lacking. We describe such a system here and use it to investigate Ty3 substrate and target sequences that influence integration.

EXPERIMENTAL PROCEDURES

Recombinant DNA Constructions—Plasmids were constructed using standard molecular biology procedures (19) unless otherwise noted. Details of plasmid constructions, plasmids and sequences of oligonucleotides used for constructions are provided in supplemental Experimental Procedures and supplemental Tables S1 and S2, respectively. Constructs were verified by DNA sequence analysis (Genewiz Inc., La Jolla, CA).

Proteins—S. cerevisiae strain BY4741 was induced to express galactose-regulated Ty3 from pDLC201 (20), and VLPs were harvested as previously described (21). Triple fusion protein (Bref1<sub>1–385</sub>-TBP<sub>61–240</sub>-Bdp1<sub>439–590</sub>) TFP) was expressed in bacterial strain Rosetta (DE3) pLysS (EMD Biosciences, San Diego, CA) and was purified essentially as described (22).

Reconstituted Ty3 IN (23) was cloned to allow expression of a C-terminal His<sub>6</sub>-tagged protein under control of the lac promoter (pKN2412). Expression was induced in Rosetta (DE3) pLysS according to standard procedures. Extracts were enriched for IN by affinity chromatography using His60 Ni Superflow. IN was further purified using anion exchange chromatography over DEAE Sephadex A-25. Details of protein purifications are provided in supplemental Experimental Procedures.

In Vitro Integration and Strand-Transfer Assays—In vitro integration using VLPs was performed as described previously (24). Either TFIIIB or TFP were mixed with target plasmids on ice for 30 min before VLPs were added, and samples were incubated at 16 °C for 15 min. Strand-transfer reactions were performed in buffer R (20 mM HEPES pH 7.5, 70 mM NaCl, 0.1% Nonidet P-40, 7.5% DMSO, 5 mM DTT) supplemented with MgCl<sub>2</sub> or MnCl<sub>2</sub> cofactors. Generally, samples contained 50 fmol of target plasmid, 250 fmol of duplex DNA, 250 fmol of TFP, and 1000 fmol of IN in a total volume of 40 µl. Reactions were incubated at 24 °C for 1 h, and DNA products were extracted as described previously (24).

PCR was performed to amplify fragments diagnostic of strand transfer. For VLP integrations, one tenth of the DNA products were combined with primers 242 and 411, which anneal within the SNR6 gene and at the downstream end of the internal domain of Ty3, respectively (25). In the PCR reactions monitoring strand-transfer products of duplex DNA substrates, primer HH1707, which anneals at the first half of the DNA substrates, was substituted for primer 411. Control PCR reactions amplified a segment of the target plasmid. Products were resolved by electrophoresis on non-denaturing 8% polyacrylamide gel or 1.5% agarose gel and visualized by staining with ethidium bromide. To determine strand-transfer sites, DNA fragments were extracted from the gel, cloned into pCR2.1 and sequenced.

Protein DNA Binding Assay—A 57-bp <sup>32</sup>P-labeled TATA-containing DNA probe was labeled, and EMSA was performed as described previously (22).

RESULTS AND DISCUSSION

A Synthetic Brf1 and TBP Fusion Protein Supports Ty3 Position-specific Integration—SNR6 is transcribed by Pol III, but is distinguished in yeast from some other Pol III templates by the presence of an upstream TATA box. TFIIIB composed of TBP, Bdp1, and Brf1 functions to dock Pol III and enhance duplex opening at the position of transcription initiation (26). In vitro TFIIIB binds DNA via interactions between TBP and the SNR6 TATA element and these interactions are sufficient to support TFIIIC-independent transcription initiation (27–28). On a template containing heteroduplex DNA at the transcription initiation site, Brf1 and TBP alone are sufficient to support transcription initiation (29). Function of TFIIIB subunits Brf1 and TBP in transcription initiation can be substituted by a structure-based fusion of the conserved domain of TBP flanked by segments of Brf1 (Bref1<sub>1–382</sub>-TBP<sub>61–240</sub>-Bref1<sub>439–590</sub>) referred to as TFP (22).

A particulate fraction containing Ty3 VLPs isolated from yeast extracts by sucrose gradient centrifugation can provide active IN and substrate cDNA (30). Ty3 VLP-mediated cDNA strand transfer differs from Pol III transcription initiation in that it can be targeted by Brf1 and TBP without introduction of heteroduplex DNA at the TSS (24). However, the requirement of Ty3 for TFIIIB and TFIIIC for integration at most tRNAs and for TFIIIB or even Brf1 and TBP at SNR6 complicates identification of interactions key to targeting. In order to better define the activities required for Ty3 strand-transfer targeting, we examined whether TFP could replace TBP and Brf1 as was found for Pol III transcription. The TATA box upstream of SNR6 can bind TBP in either orientation and thus mediate bidirectional transcription initiation at upstream (SNR6 distal) and downstream (SNR6 proximal) sites (31). A related variant on plasmid pLY1855 supports Ty3 integration at both initiation sites (25). This target plasmid was combined with bacterially-expressed TFIIIB (TBP, Brf1, and Bdp1) or TFP and Ty3 VLPs were added as the source of integration activity and cDNA (25). Strand transfer was assayed using a PCR primed by cDNA- and target plasmid-specific oligonucleotides (Fig. 1A). Products consistent with TFIIIB bound to the TATA box in each orientation were observed in positive control reactions containing TFIIIB and VLPs (Fig. 1B, lane 1) (25) and in a test reaction containing TFP (Fig. 1B, lane 2), but not in reactions containing only VLPs (Fig. 1B, lane 3). Therefore, non-conserved TBP residues 1–60 and Brf1 region 383–438, which contains HR I and the HRI-II spacer, both of which are lacking in TFP, are dispensable for targeting Ty3 strand transfer to Pol III TSS.

Recombinant IN Is Sufficient to Mediate Position-specific Integration—A remaining major limitation in defining the Ty3 components required for targeting was the requirement for a complex VLP fraction as the source of both IN and cDNA. Although it might be anticipated that IN would directly mediate specificity, recent findings in the retrovirus system indicate that domains within some retroviral Gag proteins have the capacity to influence integration patterns (2). A system was therefore developed in which recombinant IN and duplex oligonucleotides were substituted for VLPs. These strategies were previously used to reconstitute the retroviral strand-transfer
In Vitro Targeting of Strand Transfer by a Retroelement Integrase

Ty3 IN Strand Transfer Is Sensitive to Mutations in the cDNA Inverted Repeat Dinucleotide—Terminal IR are a signature feature of integrated transposons and retroviruses with TG/CA being virtually universally conserved. Upstream of the conserved dinucleotide the two ends can have distinct sequences and in vitro evolution of IN substrates has shown that additional variation is possible in the absence of requirements for replication (33). As discussed above, in the cDNA the IR copies are flanked on the outside ends by 2 “extra” bp which are removed during integration. Ty3 has a terminal 8-bp IR and 2 extra bp (plus strand, 5′-gaTGTTGTAT-3′… ATACAA-CAcc-3′). U5 oligonucleotide substrates substituted in the outside ends of the IR (CA, wt; TA, CG, TG, mutants) and a duplex oligonucleotide in which the terminal Ty3 sequence was randomized, were assayed for strand transfer (Fig. 2C). This assay showed little difference in activity among reactions using blunt or pre-processed substrates with IR sequences ending in wt CA or mutant TA (Fig. 2C, lanes 1–4). Strand-transfer products were not generated from the randomized oligonucleotide substrate (Fig. 2C, lane 9). Significantly less strand transfer was observed for processed and blunt substrates with IR ending in G, rather than wt A (Fig. 2C, lanes 5–8). In addition, among the latter templates, more strand transfer was observed for preprocessed substrates indicating that processing was sensitive to mutations of the terminal “A” (Fig. 2C, lanes 5–8). In the case of in vitro relative rate assays of HIV-1 and Ty1 and Tfp retrotransposon IN proteins, 3′-end processing of the two extra nt was blocked by mutations in the IR terminal “A” and was greatly reduced by changes in the conserved penultimate IR “C” (34) for Ty1 (35) and Tfp (36). Although these strand-transfer assays combined with PCR detection are unlikely to be as sensitive to perturbation as real time enzymatic assays, they showed that Ty3 IN activity is sensitive to changes in the terminal IR.

TFP Mediates IN Association with Target DNA—Retroviral IN proteins display robust in vitro strand-transfer activity in the absence of host targeting factors. In the case of Ty3 IN, strand-transfer assays did not show evidence of a default nonspecific pathway. Nonetheless, this activity would yield more diffuse products in our assay and therefore be more difficult to detect than specific strand transfer. Therefore, the ability of IN to

and plasmid target. The reactions including wt IN generated fragments of the size expected for Ty3 strand transfer at the divergent TSS (Fig. 2B, lane 4; Fig. 2C, lanes 1 and 2); the D22E/E261D mutant IN failed to generate these products (data not shown). Sequence analysis of four independent reactions identified eleven distinct joints of targeted strand transfers. The majority of joints were distributed within one or two nt of the TSS on the template strand or offset upstream by five nt on the nontemplate strand (Fig. 2D). A similar amount of product was generated in reactions using unprocessed duplexes (Fig. 2C, compare lanes 1 and 2). In addition, sequence analysis of strand-transfer products of the blunt substrate showed that the junction occurred at the terminal CA, so that strand transfer was preceded by removal of two nt from the 3′-end of the duplex (data not shown). These assays demonstrated for the first time that Ty3 IN is the sole Ty3 protein required to process 3′ extra nts and target strand transfer to the Pol III transcription initiation site.

Ty3 IN Strand Transfer Is Sensitive to Mutations in the cDNA Inverted Repeat Dinucleotide—Terminal IR are a signature feature of integrated transposons and retroviruses with TG/CA being virtually universally conserved. Upstream of the conserved dinucleotide the two ends can have distinct sequences and in vitro evolution of IN substrates has shown that additional variation is possible in the absence of requirements for replication (33). As discussed above, in the cDNA the IR copies are flanked on the outside ends by 2 “extra” bp which are removed during integration. Ty3 has a terminal 8-bp IR and 2 extra bp (plus strand, 5′-gaTGTTGTAT-3′… ATACAA-CAcc-3′). U5 oligonucleotide substrates substituted in the outside ends of the IR (CA, wt; TA, CG, TG, mutants) and a duplex oligonucleotide in which the terminal Ty3 sequence was randomized, were assayed for strand transfer (Fig. 2C). This assay showed little difference in activity among reactions using blunt or pre-processed substrates with IR sequences ending in wt CA or mutant TA (Fig. 2C, lanes 1–4). Strand-transfer products were not generated from the randomized oligonucleotide substrate (Fig. 2C, lane 9). Significantly less strand transfer was observed for processed and blunt substrates with IR ending in G, rather than wt A (Fig. 2C, lanes 5–8). In addition, among the latter templates, more strand transfer was observed for preprocessed substrates indicating that processing was sensitive to mutations of the terminal “A” (Fig. 2C, lanes 5–8). In the case of in vitro relative rate assays of HIV-1 and Ty1 and Tfp retrotransposon IN proteins, 3′-end processing of the two extra nt was blocked by mutations in the IR terminal “A” and was greatly reduced by changes in the conserved penultimate IR “C” (34) for Ty1 (35) and Tfp (36). Although these strand-transfer assays combined with PCR detection are unlikely to be as sensitive to perturbation as real time enzymatic assays, they showed that Ty3 IN activity is sensitive to changes in the terminal IR.

TFP Mediates IN Association with Target DNA—Retroviral IN proteins display robust in vitro strand-transfer activity in the absence of host targeting factors. In the case of Ty3 IN, strand-transfer assays did not show evidence of a default nonspecific pathway. Nonetheless, this activity would yield more diffuse products in our assay and therefore be more difficult to detect than specific strand transfer. Therefore, the ability of IN to
interact with target DNA was reinvestigated using a more direct assay. A 57-bp duplex oligonucleotide DNA containing the SNR6 TATA element was used to represent the target DNA. An identical duplex was previously used to measure binding of TFP specifically to TATA-containing DNA (22). Over a range of IN concentrations, no interaction between IN and the target DNA was observed (Fig. 3, left panel). In contrast, as reported previously, addition of TFP alone retarded mobility of the SNR6 target duplex (22). In the presence of TFP, supershifting of the TATA-containing duplex was proportional to the amount of IN (Fig. 3, middle panel). However, this interaction was weak for both wt IN and a catalytic site mutant (data not shown). Overall, these results support a model in which the Ty3 intasome interactions with Pol III promoters is mediated by direct interaction of Ty3 IN with Brf1 and TBP components of TFIIIB. This model is similar to what has been proposed for targeting of Ty5 (37) and Tf1 (15) integration by IN tethering to target-bound proteins.

In the Presence of MnCl₂, Strand Transfer Is TFP-independent and Sequence Specific—In vitro substitution of the natural MgCl₂ metal cofactor with MnCl₂ in the case of HIV-1 IN reduces specificity for cDNA termini (34) and enhances activity in disintegration assays (38). To test the effect of MnCl₂ on the association of Ty3 IN with its target, MgCl₂ was either supplemented or substituted with MnCl₂ in the strand-transfer reactions. PCR analysis of products of MnCl₂-containing substrates showed surprisingly that strand transfer was dependent upon IN, but independent of TFP (Fig. 4A). Strand transfer was not observed for the randomized oligonucleotide substrate in the presence of MnCl₂, indicating that it required specific interactions with IN (data not shown). Reactions containing VLPs showed only a low level of non-targeted products in the presence of MnCl₂ (supplemental Fig. S1). In the presence of MnCl₂, TFP shifted the TATA-containing probe indicating that MnCl₂ does not produce TFP-independent strand transfer by disruption of TFP binding (Fig. 3, right panel). However, the IN supershift was no longer observed, suggesting that the presence of MnCl₂ affected the interaction between TFP and IN.

The PCR amplicon from products generated in the presence of MnCl₂ concentrations greater than 10 mM in the presence or
absence of TFP was ∼300 bp (Fig. 4, A and B). Experiments were performed in which MnCl₂ or MgCl₂ was increased in the absence or presence of the other metal cation and the products were amplified using PCR (Fig. 4B and data not shown). In high MgCl₂ and low MnCl₂ bands representing products of strand transfers flanking the TFP binding site were observed as previously described. However increasing MnCl₂ correlated with increasing amounts of higher molecular weight products including a major product of about 300 bp and decreasing amounts of lower molecular weight products (Fig. 4B). Since these products were clearly discrete from previously observed targeted strand-transfer products, products of three independent reactions were cloned and submitted for sequencing. This analysis showed strand transfer mainly within a small region. Among the six sites revealed by sequencing, four occurred within a 5-bp region from −231 to −226 upstream of SNR6 and the others occurred at positions −285 and −125 (Fig. 4C). One possibility was that strand transfer at a secondary TFP-binding site was observed as previous described. Instead, the sequence (5’-ACAACATA-3’) resembling the terminal IR of Ty3 (5’-ACAACACA-3’/3′-TGTGT-3’) was identified between −213 and −205 upstream of SNR6. The four clustered positions of strand transfer occurred 13 to 18 nt upstream of the 5’-end of this sequence (Fig. 4C).

The strand-transfer products recovered in the vicinity of plasmid sequences resembling the Ty3 IR suggested that IN might confer sequence specificity to strand transfer under some conditions. To directly test whether Ty3 IN targeted Ty3 IR-like sequences, a plasmid containing an isolated Ty3 LTR truncated at the downstream end to remove one IR (pXQ2889) was used as a target (Fig. 4D). Strand-transfer assays were performed using MnCl₂ as the cation and the preprocessed U5 oligonucleotide duplex substrate. PCR products of this reaction showed dominant fragments of 300 bp (Fig. 4D, lane 1). The mixed PCR products were cloned and sequenced of six clones were determined. This analysis showed strand-transfer joints at positions −13, −9, and −7 relative to the outside end of the target Ty3 LTR (5’-TGTTGTAT-3’). To assess the distribution of target sites more completely, cloned strand-transfer products at these positions were used as templates to obtain ³²P-labeled markers. Migration of these markers was compared with that of products of an independent strand-transfer reaction. Comparison of the distribution of PCR fragments templated by products of the total strand-trans-
In vitro Targeting of Strand Transfer by a Retroelement Integrase

In the presence of Mg²⁺ chromosomal context would lie outside of Ty3.

Deletion of the strand-transfer reaction on the Ty3 IR, the target plasmid was modified in the Ty3 IR from TGGTGATT to TCACGATT to produce plasmid pXQ3673 (Fig. 4D). In contrast to PCR templated by the reaction using the wt IR target, PCR of the reaction containing the mutated IR failed to generate a product (Fig. 4D, lane 2). Control PCR reactions monitoring DNA recovery showed no difference in plasmid recovery between the two sets of samples (data not shown). Although it appears that in MgCl₂, in the absence of a targeting factor, strand transfer does not occur or is extremely inefficient, this may be because the PCR assay is less sensitive to detection of highly distributed products. In the presence of MnCl₂, strand transfer was independent of TFP and sensitive to detection of highly distributed products. In the extremely inefficient, this may be because the PCR assay is less sensitive to detection of highly distributed products. In the presence of MnCl₂, strand transfer was independent of TFP and concentrated near Ty3 IR-like sequences. Thus, IN strand transfer activity per se does not depend upon the presence of a targeting transcription factor. U5 strand transfer was only observed upstream of the Ty3 IR. This is consistent with asymmetric targeting by the 8-bp sequence to regions which in a chromosomal context would lie outside of Ty3.

The experiments in which the position of TFP was shifted by IN in the presence of Mg²⁺ but not in the presence of Mn²⁺, together with the redirection of strand transfer in the presence of MnCl₂, suggested that IN interacts directly with IR-containing DNA in the presence of MnCl₂. However, gel shift experiments similar to those which detected weak TFP-mediated IN association with TATA-containing target probe failed to identify detectable binding to IR-containing 50-mers (data not shown).

The possibility that the presence of MnCl₂ enhances weak sequence-specific interactions is intriguing. The crystal structure of the primate foamy virus intasome (39) showed a dimer of dimers with the catalytic site at the dimer-dimer interface; residues interacting with the donor IR mapped to the catalytic core and C-terminal domains of the interface. If we assume a similar structure for Ty3 IN, outer subunits might be available to participate in targeting. We speculate that in the presence of MgCl₂, they interact preferentially with the TFIIB complex, whereas in the presence of MnCl₂, this interaction is disfavored and IR-interacting residues mediate interactions (Fig. 4F).

Although this activity is interesting in terms of intasome structure-function, it may have minimal in vivo significance. In vivo, integration into Pol III initiation sites clearly dominates (40) and the concentration of MnCl₂ required for IR targeting was significantly greater than the reported physiologic concentration (41).

In summary, the involvement of multiple retroelement and host proteins and poorly-defined insertion preferences complicates elucidation of retroelement targeting. This study reconstitutes precise retroelement targeting in vitro for the first time and delimited the retroelement and host components responsible. Intriguingly, our studies showed that both protein-targeted and IR sequence-targeted modes of strand transfer can occur in vitro. We propose that outer intasome subunits not involved in strand transfer are available for target interaction.

REFERENCES

12. Devine, S. E., and Boeke, J. D. (1996) Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase III. Genes Dev. 10, 620–633

Acknowledgments—We thank B. Irwin for technical support and helpful discussions. We thank K. Nguyen for providing the IN expression plasmid. We thank G. A. Kassavetis and E. P. Geiduschek, University of California, San Diego, for providing TFIIB and TBP, BRF1, and TFP expression plasmids and for many helpful discussions.