In Vivo Retroviral Integration: Fidelity to Size of the Host DNA Duplication Might Be Reduced When Integration Occurs near Sequences Homologous to LTR Ends

Karen Moreau, Caroline Torne-Celer, Claudine Faure, Gérard Verdier, and Corinne Ronfort

Centre de Génétique Moléculaire et Cellulaire, Centre National de la Recherche Scientifique, Institut National de la Recherche Agronomique, Université Claude Bernard, 69622 Villeurbanne cedex, France

Received May 17, 2000; returned to author for revision August 7, 2000; accepted September 10, 2000

Integrated retroviral DNAs are flanked by a short duplication of target DNA whose size is virus specific and invariable. We have sequenced the junctions between an ALSV (Avian Leukemia and Sarcoma Viruses)-based vector and quail DNA from five individual proviruses. Three proviruses were flanked by the expected 6-bp duplication of host DNA, whereas the two others were flanked by a 5-bp duplication. Nucleotide sequencing of the native integration sites of these two proviruses showed that these integrations had occurred at the immediate vicinity of either a CA or a TG dinucleotide, revealing striking microhomologies between the integration sites and viral LTR ends. These results suggest that size duplication of the target DNA might be influenced by nucleotide sequence at the site of integration.

INTRODUCTION

Retroviruses are single-stranded RNA-containing viruses that replicate through a DNA intermediate in cells of diverse vertebrate origins (Luciw and Leung, 1992). Following retroviral infection of a permissive cell, the viral RNA is reverse transcribed into a DNA copy of the viral genome, which is subsequently integrated into the host genomic DNA. The integrated viral DNA is transcribed and the resulting RNA either is used for translation of viral proteins or serves as the viral genome (Luciw and Leung, 1992).

Integration is carried out by the retroviral integrase (IN) protein (for recent reviews on the integration mechanism, see Brown, 1997; Hansen et al., 1998). IN recognizes the blunt ends of the viral linear DNA whose juxtaposition forms the attachment site (att) for integration into host chromosomal DNA. Integration proceeds in three steps. In the first step, 3′-processing, the linear blunt-ended DNA duplex with terminal sequences 5′-NNTG...CANN-3′ is nicked by IN on the 3′ of the CA to produce new 3′-hydroxyl ends (CAOH-3′) that are recessed by 2 nt. In the second step, strand transfer, the viral 3′-hydroxyl ends are coupled to phosphate groups in opposite strands of the target DNA, a few base pairs apart. Subsequently, in the third step, the gaps in target DNAs are filled in, probably by cellular enzymes (Daniel et al., 1999). As a result, proviruses have the TG and CA dinucleotides at the 5′ and 3′ boundaries, respectively, and are flanked by short direct repeats of the target DNA (Brown, 1997; Hansen et al., 1998).

The length of the target DNA duplication is virus specific, corresponding to 4 bp for MuLV (Murine Leukemia Virus), 5 bp for HIV (Human Immunodeficiency Virus), and 6 bp for ALSV (Avian Leukemia and Sarcoma Virus) (Varmus and Brown, 1989; Brown, 1997). In vivo, fidelity to a specific duplication size appears to be relatively strong (Varmus and Brown, 1989). However, when purified IN and reconstituted in vitro systems are used, fidelity to size is reduced (Aiyar et al., 1996; Fitzgerald and Grandgenett, 1994).

We recently developed an ALSV-based vector (NP3Catt; Fig. 1A) that carries an additional copy of the att sequence and showed that some proviruses were flanked by the additional att sequence at one extremity and by the viral LTR at the other extremity (Torne-Celer et al., manuscript in preparation). In the course of this work, we also isolated proviruses that arose from insertion using the natural termini of LTRs, whose structure is depicted in Fig. 1B. The molecular cloning of proviral DNA is described in detail in the above-mentioned report (Torne-Celer et al., in preparation). Briefly, QT6 cells were infected with NP3Catt helper-free viral stocks and selected with puromycin. Puro-resistant (puroR) clones were isolated and proviruses were analyzed by Southern blot and PCR techniques. Five LTR-integrated proviruses were identified. All of them underwent a deletion of the neo gene, bringing together sequences in the LTR or leader region and internal att sequences. Recombinations involved short stretches of sequence identity (2 to 4 nt in length) and were probably the result of recombina-
tions that occurred during the reverse transcription step (Torne-Celer et al., manuscript in preparation). We characterized the cell–virus DNA junctions of these proviruses by inverse PCR amplifications (Ochman et al., 1988) and DNA sequence analysis (Torne-Celer et al., manuscript in preparation).

RESULTS AND DISCUSSION

This evaluation revealed that these proviruses were derived by correct integration, since (i) they sustained a loss of 2 nt at each viral end and (ii) they were flanked by direct repeats of host flanking DNA. In three of the five integrants analyzed (proviruses att5, att6, and att8), we were able to identify unambiguously the expected 6-bp direct repeat of host DNA (Fig. 1B). Two other proviruses (att2 and att10) were flanked by a 5-bp repeat rather than by a 6-bp duplication of target DNA. However, analysis of both viral and cellular sequences at junctions revealed that a TTCA (provirus att2) and a TGT (provirus att10) were present on both viral and flanking cellular sequences. If these bases were taken into account, this would lead to duplication sizes of either 9 or 8 bp, respectively.

As analysis of the host–viral junctions did not allow an unambiguous determination of the host DNA duplication size of proviruses att2 and att10, we sequenced the native integration sites of these proviruses, as well as the integration site of proviruses att5 and att6 as controls. Primers were selected on both the left and right host DNA sequences and used for a PCR amplification on QT6 cells DNA as template. Resulting PCR products were then directly sequenced. Analysis of the integration sites of proviruses att5 and att6 confirmed the 6-bp duplication of target sequence, since this sequence was

![Diagram of LTR-integrated NP3Catt puroR proviruses](image-url)

**FIG. 1.** Genetic structure of LTR-integrated NP3Catt puroR proviruses. (A) Map of the retroviral vector NP3Catt. The large boxes are LTRs, marker genes (neo and puro selectable genes), and SV40 sequences (promoter [pro] and polyadenylation [pA] signal). The small boxes represent part of the gag and env genes (δG and δE) and the thin lines represent viral noncoding sequences (leader, J region from AEV [Avian Erythroblastosis Virus] and 3' noncoding sequences; ¥, psi [packaging signal]). SD and SA: splice donor and splice acceptor sites, respectively. An additional att sequence (containing 46 bp of U5 end and 45 bp of U3 end) is depicted by a hatched box. (B) DNA sequences at the host–viral junctions of LTR-integrated puroR proviruses. PuroR clones were isolated following infection of QT6 cells with NP3Catt viral stocks and selection with puromycin drug. Detailed molecular analysis of the proviruses and the strategy of sequencing junction DNA are described elsewhere (Torne-Celer et al., in preparation). Five proviruses were found to be LTR-integrated. On the top is depicted such an LTR-integrated puroR provirus flanked by host chromosomal DNA (thick lines). These proviruses suffered a deletion of neo gene and surrounding sequences, bringing together nucleotides from U5 or pbs regions (represented in dotted lines) and the internal att sequence. Residual U3 sequences of the internal att sequence (ΔU3) varied in size from 0 (provirus att10) to 48 bp (provirus att5) (Torne-Celer et al., in preparation). Below are represented viral extremities of blunt-ended precursor of integration showing 2 nt (in italics) removed during the integration process. The viral terminal CA and TG dinucleotides are underlined. For each viral end, more than 100 bp of flanking DNA sequences were determined but only 20 bp from each end are shown. Cellular sequences duplicated on both sides are represented in gray boxes. When examining both viral and cellular sequences, duplications could be considered longer in proviruses att2 and att10 (nt in white boxes). GenBank accession numbers for full-length sequences are: AY004217 (att2), AY004218 (att5), AY004219 (att6), AY004220 (att8), AY004221 (att10).
found in only one copy before integration (data not shown). Analysis of the integration sites of proviruses att2 and att10 (Fig. 2A) (i) confirmed the 5-bp-long duplication and (ii) revealed the presence of a TTCA (provirus att2) and of a TGT (provirus att10) in the target DNA before integration. In retroviral DNAs analyzed so far, the viral ends were systematically cut after the conserved CA and TG dinucleotides. Assuming that similar rules were applied to proviruses att2 and att10, we conclude that integration of proviruses att2 and att10 occurred in a region with a small stretch of homologies with viral ends (TTCA for provirus att2 and TGT for provirus att10). Finally, the direct cellular repeat flanking proviruses att2 and att10 is 5 nt long (Fig. 2A).

To our knowledge, no such incorrect duplications have been previously reported during avian retroviral integration in vivo. However, there are at least two precedents in the MuLV system. While Mo-MuLV viruses generate a 4-bp duplication of target DNA, two Mo-MuLV mutants (in592-2 and dl594) were flanked by a 1- and 5-bp duplication, respectively (Colicelli and Goff, 1985, 1988). Similarly, of the three examined, two AKR-MuLV DNAs (AKR-623 and AKR-614) generated a 3- and 5-bp duplication (Van Beveren et al., 1982). Clearly, the four integrants showing duplications smaller than expected (proviruses att2 and att10, in592-2, AKR-623) (Figs. 2A and 2B) exhibit either a CA dinucleotide on the left side of the cellular duplicated sequences or a TG dinucleotide on the right side of the duplicated sequences, precisely at the point of nucleophilic attack by the viral CAOH-3’ end.

It is difficult to assess the significance of altered host-site duplication with respect to the CA or TG connection of the target sequence. However, we speculate that during the nucleophilic attack of the target DNA by the CA-3’ end, the hybridization of the viral CA to the cellular TG led to a mispositioning of the integrase and, subsequently, to a smaller duplication size.

Despite the great number of sites sequenced for HIV (Stevens and Griffith, 1996; Carteau et al., 1998) and Human T-cell Leukemia Virus type 1 (Leclercq et al., 2000), the influence of microhomologies on duplication size has never been observed. However, as similar observations were previously done with MuLV viruses, we do not believe these observations of unusual duplication size are the result of specific particularities of our system. Although the sample size is small, we note that the frequency of unexpected duplications is high (two out of five proviruses examined). Therefore, we cannot yet say whether the presence of internal att sequences increases the frequency of these peculiar events.

IN proteins are known to integrate viral DNA by a nonhomologous recombination mechanism (Brown, 1997). However, it has been shown that the sequence of the host site can influence the choice of the target site (Fitzgerald and Grandgenett, 1994). The results presented here strongly suggest that the sequences of the host site sometimes may also influence the fidelity to size of the host DNA duplication generated at the site of integration.

ACKNOWLEDGMENTS

This work was supported by research grants from the Centre National de la Recherche Scientifique (CNRS), the Institut National de la Recherche Agronomique (INRA), and the Association pour la Recherche sur le Cancer (ARC). We thank Transgene SA (Strasbourg), the INRA, the Ligue Nationale contre le Cancer, and the Ministere de la Recherche for fellowships (C.T.C. and K.M.).

REFERENCES


