JOURNAL OF VIROLOGY, Nov. 1990, p. 5645–5648
0022-538X/90/115645-04$02.00/0

Vol. 64, No. 11

Sequence Requirements for Integration of Moloney Murine Leukemia Virus DNA In Vitro

FREDERIC D. BUSHMAN* AND ROBERT CRAIGIE

Laboratory of Molecular Biology, National Institute of Diabetes, and Digestive and Kidney Diseases, Building 2, Room 218, Bethesda, Maryland 20892

Received 17 May 1990/Accepted 7 August 1990

Normal replication of Moloney murine leukemia virus (MoMLV) requires the integration of a DNA copy of the viral RNA genome into a chromosome of the host. In this work, we characterize the DNA sequences at the ends of the linear proviral precursor that are required for integration in the presence of MoMLV integration protein in vitro. We found that nine bases of MoMLV DNA at each end of a linear model substrate were sufficient for near-maximal levels of integration and that four bases of MoMLV DNA at each end were sufficient for low levels of correct integration. We also found that a 3′-terminal A residue was preferred for integration. We infer from the limited DNA sequence requirements for integration that factors in addition to DNA sequence direct integration protein to act at the ends of the viral DNA.

To replicate efficiently, a retrovirus must integrate a DNA copy of its RNA genome into a chromosome of the host. For recent reviews of retroviral replication and the role of integration, see references 7 and 14.

Studies of integration in the Moloney murine leukemia virus (MoMLV) system indicate that integration involves the following steps. The viral integration (IN) protein initially prepares the linear DNA product of reverse transcription for integration by specifically cleaving the flush DNA ends to produce recessed 3′ ends (1a, 3a, 10). A similar cleavage activity has also been demonstrated for the IN protein of avian sarcoma-leukemia virus (8). Next, IN protein joins these 3′ ends to the 5′ ends of a double-stranded break in the target DNA (1a, 3a, 6). The breaks on each strand of the target DNA are apparently offset by 4 bp, because MoMLV proviruses are found to be flanked by 4-bp duplications of target DNA (13). Repair of the single-stranded DNA gaps in the integration intermediate completes the integration of viral DNA.

Analysis of mutant derivatives of MoMLV identify two viral functions involved in integration, IN protein and DNA sequences at the ends of the linear viral DNA. Viruses mutant in either of these functions have similar phenotypes: normally integrated viral DNA does not accumulate, and replication is impaired (2, 4, 12). This finding is consistent with the view that IN protein binds to the viral DNA termini. Terminal DNA sequences important for integration in vivo appear to overlap 13-bp inverted repeats (IRs) located at each end of the linear viral DNA. Evidence for this view comes from an analysis of the phenotypes of mutants containing deletions of one IR. A mutant containing an internal deletion that retains only the terminal 13-bp IR is competent for integration (cited in reference 10), whereas mutants missing 2 or 8 terminal bp from one IR are impaired in integration (2).

We have used an in vitro integration system to characterize the terminal sequences of MoMLV DNA required in the integration reaction. Our assay scores the insertion of model integration substrates into a lambda DNA target (5). Integration activity is provided by MoMLV IN protein overexpressed in insect cells and partially purified as described elsewhere (3a). Our standard model substrate, termed miniMoMLV, is a linear DNA molecule containing at each end 21 of the terminal 22 bases of the linear proviral DNA precursor. Position 1 of miniMoMLV was changed from the viral A to T to allow linearization of plasmid pMK471 with *Nde*I to produce miniMoMLV. Mutant viruses containing a variety of alterations in length and sequence in the first two positions are competent for replication in vivo, and this particular change is known to be tolerated (2, 3). The 3′ ends of miniMoMLV are recessed, thereby bypassing the need for the long-terminal-repeat-specific IN nuclease. To assay integration, partially purified IN protein, miniMoMLV, and lambda DNA are incubated in a suitable reaction mixture, and integration products are deproteinized and packaged into lambda bacteriophage heads in vitro. Integration reactions were carried out exactly as described in reference 5 except that the extract of disrupted virion particles was replaced with 4 pmol of partially purified IN protein. The resulting phage particles are used to infect an *Escherichia coli* strain containing a lambda prophage. Lambda cannot grow on the lysogenic strain, but phage genomes containing integrated miniMoMLV can replicate as plasmids because miniMoMLV contains the pBR322 origin of replication. Because miniMoMLV also contains genes conferring resistance to ampicillin and tetracycline, cells containing recombinant phage genomes can form colonies on selective plates.

To examine the viral DNA sequences required for integration, we constructed and assayed the set of modified substrates shown in Fig. 1. MiniMoMLV Δ9 contains nine bases matching miniMoMLV DNA at each end, and miniMoMLV Δ6/7 contains six of seven bases matching miniMoMLV DNA at each end (counting from the end of the unpaired 5′ extension of the linear DNA). We also tested two other model substrate DNAs: a linear DNA containing four bases of miniMoMLV sequences at each end (pBR/Nde 1), and a linear DNA containing no terminal miniMoMLV sequences and *Eco*RI overhanging ends (pBR/EcoR 1). Each modified substrate also contains additional matches to miniMoMLV contributed by vector DNA sequences. As with miniMoMLV, the 3′ ends of these substrates are recessed, thereby bypassing the need for the long-terminal-repeat-specific nuclease activity of IN protein.

* Corresponding author.

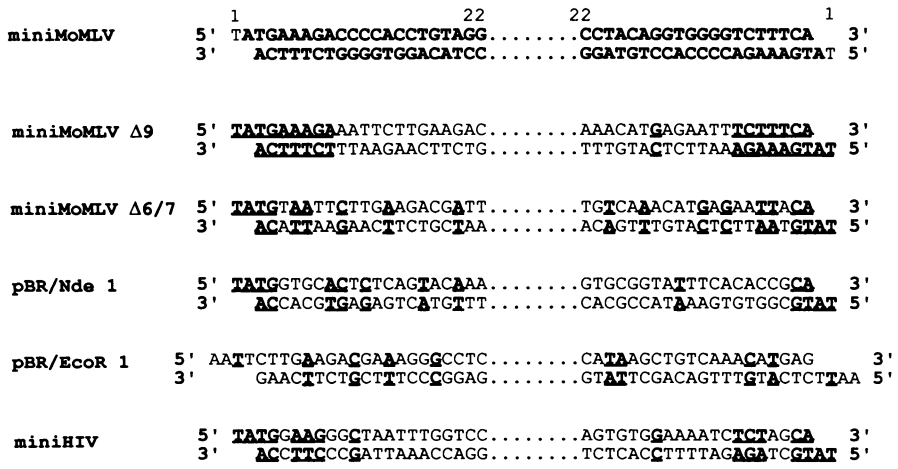

FIG. 1. Model integration substrates. The numbering scheme of the miniMoMLV bases is as indicated. Matches between the MoMLV DNA termini and miniMoMLV are shown in boldface. Matches between miniMoMLV and the other model substrates are shown in boldface and underlined. The 22 bases at each end of miniMoMLV match the predicted sequence of the U3 end of the unintegrated MoMLV DNA except at position 1, which was changed to allow linearization of plasmid pMK471 with *Nde*I to produce miniMoMLV. MiniMoMLV Δ9 and Δ6/7 were constructed by ligating double-stranded oligonucleotides flanked by *Eco*RI cohesive ends into the *Eco*RI site of pMK468 (5, 9). Purified plasmids were linearized with *Nde*I. pBR/Nde 1 and pBR/EcoR 1 were constructed by linearizing pBR322 with *Nde*I and *Eco*RI, respectively.

The efficiencies of integration of these model substrates were compared by adding 0.1 μg of each substrate to a standard integration reaction and comparing the numbers of recombinant-containing colonies obtained (Table 1). MiniMoMLV Δ9 and miniMoMLV Δ6/7 yielded 77 and 43% as many colonies as did the standard miniMoMLV. PBR/Nde 1 yielded only 2% as many colonies as did miniMoMLV, and pBR/EcoR 1 yielded 0.4% as many colonies as did miniMoMLV.

The structures of several integration products generated in each model substrate were analyzed by sequencing reactions containing the junctions between the model substrate DNA and the lambda target DNA (Table 1 and Fig. 2). As with MoMLV integration in vivo, the sites of integration in vitro showed no obvious sequence specificity. Twenty-one of 24 of the recombinants sequenced contained the 4-bp duplications of target DNA characteristic of MoMLV DNA integration in vivo. The three products lacking 4-bp duplications all contained deletions of at least 1 kb of target DNA at the point of insertion. The sequence of one such event is shown in Fig. 2a. Such products might have been formed by

TABLE 1. Results of integration assays containing various model substrates[a]

Substrate	No. of colonies	Relative efficiency (%)	No. of correct junctions at both ends/no. sequenced
MiniMoMLV	13,490	100	4/5
MiniMoMLV Δ9	10,324	77	3/4
MiniMoMLV Δ6/7	5,766	43	3/4
pBR/Nde 1	248	2	3/6
pBR/EcoR 1	51	0.4	0/5

[a] The sum of results from two independent experiments is presented. Incorrect junctions derived from reactions containing miniMoMLV, miniMoMLV Δ6/7, and pBR/Nde 1 each contained one example of a large target deletion with no 4-bp duplication of target sequences. The remainder of the aberrant products have base deletions or additions at the end(s) of the substrate and 4-bp duplications.

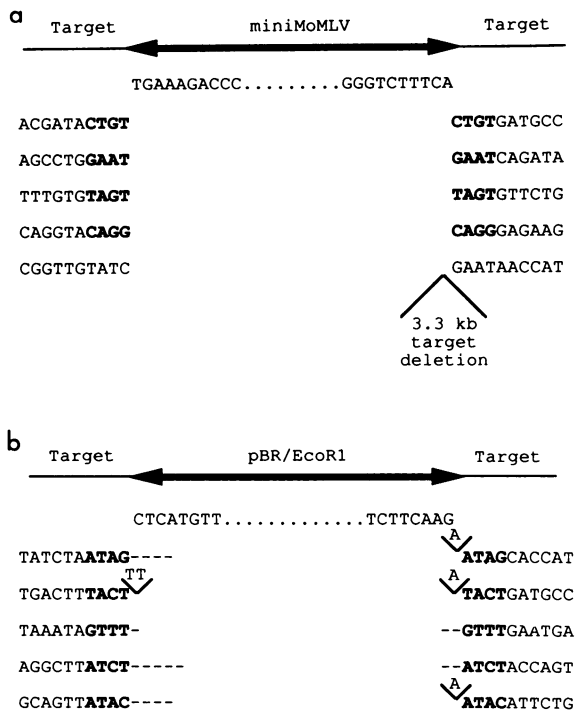

FIG. 2. Sequences of junctions between model substrate DNAs and target DNA in integration products. Sequences of the ends of the model substrate are shown below the heavy arrows. Target sequences are shown to either side of the model substrate. Target duplications are shown in bold letters. A target deletion is as indicated below the sequence; base insertions are indicated above the sequences. Each base deleted in the pBR/EcoR 1 DNA is indicated with a dash. In a few cases, an ambiguity arose because the target DNA sequence allowed more than one interpretation of the sequence changes resulting from integration. In such cases, we chose the interpretation that preserved the 4-bp duplication. DNA products of integration were sequenced by the method of Sanger et al. (11), using the primers described in reference 5.

uncoupled attack by the two ends of the model substrate on target DNA or by recombination between two lambda genomes containing independent insertions.

We found modifications of the substrate DNA, including deletions of the substrate DNA and base additions, in some of our integration products (Table 1). These modifications were all found at the junctions between model substrate and target DNAs, as would be expected if the ends of the model substrate were modified before integration. The DNA-modifying activities were apparently contained in NIH 3T3 cell extracts that were added to the reactions to increase the efficiency of integration (5). Recovery of colonies was stimulated about 50-fold by NIH 3T3 extracts, as was previously reported for similar reactions with disrupted virions as the source of IN protein. The mechanism of this stimulation remains unclear. Products of reactions lacking these extracts do not contain base additions and substrate deletions (data not shown). Substrate modifications were most common in products of integration of our least efficient substrate, pBR/EcoR 1, in which every end that appeared in the DNA products analyzed was modified. No integration of pBR/EcoR 1 was detected in reactions lacking the NIH 3T3 extracts (data not shown). In 9 of 10 cases, these modifications resulted in the placement of an A residue at the 3′ end that became joined to target DNA (Fig. 2b). These pBR/EcoR 1 molecules appear to have been integrated normally after modification, since each product contained the normal 4-bp duplication of target DNA.

Analysis of the integration efficiencies of our various model substrates supports the following conclusions. (i) The finding that miniMoMLV Δ9 is integrated about as efficiently as miniMoMLV suggests that the sequence of bases 10 to 22 is not very important in integration. (ii) MiniMoMLV Δ6/7 is integrated 43% as well as miniMoMLV and matches miniMoMLV at six of the seven terminal positions, highlighting the importance of the terminal 7 bp. PBR/Nde 1 is integrated only 2% as well as miniMoMLV and matches the terminal four bases of miniMoMLV, indicating that the sequence at positions six and seven has a strong quantitative effect on integration. (iii) An A residue at the 3′ end of the model substrate DNA is apparently favored for integration, since 9 of 10 of the ends found in integration products of pBR/EcoR 1, our least efficient substrate, were apparently modified to position an A at the junction with target DNA. We interpret the exclusive use of modified pBR/EcoR 1 molecules as reflecting the results of a competition in vitro, in which these modified pBR/EcoR 1 molecules were selected by the integration machinery from a pool containing both modified and unmodified pBR/EcoR 1 molecules.

The finding that the terminal nine bases of miniMoMLV are sufficient for near-maximal efficiency of integration explains the observation that a miniHIV model substrate DNA can be integrated by the MoMLV integration machinery (unpublished data). MiniHIV contains 30 bp of matching human immunodeficiency virus (HIV) DNA at each end (Fig. 1) and can be integrated into a DNA target in the presence of partially purified HIV IN protein (F. D. Bushman, T. Fujiwara, and R. Craigie, Science, in press). We find that miniHIV is integrated only about three times less efficiently than miniMoMLV in the presence of MoMLV IN protein. Inspection of the miniMoMLV and miniHIV sequences reveals that miniHIV matches seven of nine of the terminal bases of miniMoMLV (Fig. 1). This observation simply explains the efficient integration of miniHIV in the presence of MoMLV IN protein and further highlights the importance of the terminal nine bases for integration of MoMLV DNA.

The observation that so few bases of the IR sequence are sufficient for DNA integration in vitro, and so presumably for recognition by IN protein, raises the question of how IN protein specifically recognizes its site. This finding does not appear to be an artifact of the in vitro assay, because our results are broadly consistent with results obtained in vivo by others (2, 3; J. E. Murphy and S. P. Goff, personal communication). Taken together, results obtained in vivo and in vitro indicate that roughly 5 bp of the IR (positions 3 to 7) are sufficient for recognition by IN protein, probably too short a sequence to provide adequate discrimination on its own between viral and host DNA. Perhaps IN protein recognizes not only the base sequence of the IR but some aspect of the nearby DNA end as well (10). Another possibility, not exclusive of the first, is that IN protein is helped in recognizing its site by interactions with other proteins. In the cytoplasm, IN protein forms part of a large nucleoprotein complex containing the reverse-transcribed viral DNA and several viral proteins (1, 1b, 14). If this complex persists in the nucleus, protein-protein interactions within the complex may position IN protein near the viral DNA termini before integration. Thus, relatively little sequence specificity may be required in the interaction between IN protein and its site because IN protein may never be free to equilibrate between the viral and host DNAs.

This work was supported by the National Institutes of Health Intramural AIDS Targeted Antiviral Program. F.D.B. is a fellow of the Leukemia Society of America.

LITERATURE CITED

1. **Bowerman, B., P. O. Brown, J. M. Bishop, and H. E. Varmus.** 1989. A nucleoprotein complex mediates the integration of retroviral DNA. Genes Dev. **3**:469–478.
1a. **Brown, P. O., B. Bowerman, H. E. Varmus, and J. M. Bishop.** 1989. Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein. Proc. Natl. Acad. Sci. USA **86**:2525–2529.
1b. **Brown, P. O., B. Bowerman, H. E. Varmus, and J. M. Bishop.** 1987. Correct integration of retroviral DNA in vitro. Cell **49**: 347–356.
2. **Colicelli, J., and S. P. Goff.** 1985. Mutants and pseudorevertants of Moloney murine leukemia virus with alterations at the integration site. Cell **42**:573–580.
3. **Colicelli, J., and S. P. Goff.** 1988. Sequence and spacing requirements of a retrovirus integration site. J. Mol. Biol. **199**:47–59.
3a. **Craigie, R., T. Fujiwara, and F. D. Bushman.** 1990. The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration in vitro. Cell **62**: 829–837.
4. **Donehower, L. A., and H. E. Varmus.** 1984. A mutant murine leukemia virus with a single missense codon in *pol* is defective in a function affecting integration. Proc. Natl. Acad. Sci. USA **81**:6461–6465.
5. **Fujiwara, T., and R. Craigie.** 1989. Integration of mini-retroviral DNA: a cell-free reaction for biochemical analysis of retroviral integration. Proc. Natl. Acad. Sci. USA **86**:3065–3069.
6. **Fujiwara, T., and K. Mizuuchi.** 1988. Retroviral DNA integration: structure of an integration intermediate. Cell **54**:497–504.
7. **Grandgenett, D. P., and S. R. Mumm.** 1990. Unraveling retrovirus integration. Cell **60**:3–4.
8. **Katzman, M., R. A. Katz, A. M. Skalka, and J. Leis.** 1989. The avian retroviral integration protein cleaves the terminal sequences of linear viral DNA at the in vivo sites of integration. J. Virol. **63**:5319–5327.
9. **Maniatis, T., E. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

10. **Roth, M. J., P. L. Schwartzberg, and S. P. Goff.** 1989. Structure of the termini of DNA intermediates in the integration of retroviral DNA: dependence on IN function and terminal DNA sequence. Cell **58**:47–54.
11. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463.
12. **Schwartzberg, P., J. Colicelli, and S. P. Goff.** 1984. Construction and analysis of deletion mutations in the *pol* gene of Moloney murine leukemia virus: a new viral function required for productive infection. Cell **37**:1043–1052.
13. **Shoemaker, C., S. P. Goff, E. Gilboa, M. Paskind, S. W. Mitra, and D. Baltimore.** 1980. Structure of a cloned circular Moloney murine leukemia virus DNA molecule containing an inverted segment: implications for retrovirus integration. Proc. Natl. Acad. Sci. USA **77**:3932–3936.
14. **Varmus, H. E., and P. O. Brown.** 1989. Retroviruses, p. 53–108. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.