JOURNAL OF VIROLOGY, Nov. 1988, p. 3958–3964
0022-538X/88/113958-07$02.00/0

Vol. 62, No. 11

Analysis of Mutant Moloney Murine Leukemia Viruses Containing Linker Insertion Mutations in the 3′ Region of *pol*

LAWRENCE A. DONEHOWER

Department of Virology and Epidemiology, Baylor College of Medicine, Houston, Texas 77030

Received 6 June 1988/Accepted 26 July 1988

Twelve linker insertion mutations have been constructed in the 3′ part of the *pol* gene of Moloney murine leukemia virus. This region of the Moloney murine leukemia virus genome encodes IN or p46^{pol}, which is required for integration of the retroviral DNA into the host cell chromosome. Viral proteins synthesized by these mutants were used to pseudotype a *neo*-containing retroviral vector. Ten of twelve linker insertion mutant pseudotypes were unable to generate stable proviruses in infected mouse cells, as measured by the formation of G418-resistant colonies. Two mutants mapping at the 3′ terminus of the IN-encoding region were competent for the formation of stable vector proviruses (hundreds of G418-resistant colonies per mutant pseudotype-infected plate). Representative linker insertion mutants were also tested for the ability to synthesize viral unintegrated DNA in newly infected cells. All assayed mutants were capable of synthesizing all normal forms of viral unintegrated DNA. The structure of integrated vector proviruses generated by defective and nondefective linker insertion mutants was also analyzed. All replication-competent mutants generated normal proviruses, while the few obtainable proviruses generated by replication-defective mutants were sometimes aberrant in structure. These results argue strongly (and confirm previous data) that the IN-encoding region of *pol* does not play a significant role in DNA synthesis, but is absolutely required for the formation of normal proviral DNA.

The *pol* gene of replication-competent murine retroviruses encodes several enzymatic activities crucial to the viral life cycle. The 5′ region of *pol* encodes the viral protease, p14^{pol} or PR (24), which cleaves the *gag* and *gag-pol* polyprotein precursors into their processed mature proteins (7, 40). The viral reverse transcriptase, p80^{pol} or RT (24), contains RNA-dependent and DNA-dependent DNA polymerase activities and enables the virus to synthesize in the infected cell a full-length viral DNA with long terminal repeats (LTRs) (6, 12, 21, 36, 37, 39). In addition, RT contains an RNase H activity, which allows formation of the double-stranded viral DNA by degradation of the viral RNA when it is in an RNA-DNA duplex (12, 36, 39). The 3′ region of *pol* codes for p46^{pol} or IN (8, 23, 28, 32). This region of *pol* is required for the efficient integration of the viral DNA into the host cell chromosome (8, 29, 32).

There is biochemical and genetic evidence that IN enzymatically mediates retroviral integration. Purified polypeptides encoded by the 3′ region of *pol* in avian and murine retroviruses have exhibited DNA-binding and DNA endonuclease activity in vitro (14, 23, 27). Recent studies on the avian virus endonuclease activity indicate that it preferentially nicks single- and double-stranded viral circular DNAs at the junction between the two viral LTRs to potentially form a precursor DNA molecule for integration (5, 9, 15). The viral IN may also be responsible for staggered cleavage of the chromosomal target DNA to generate short duplications of 4, 5, or 6 base pairs (bp) that flank the provirus (although there is no direct evidence for this activity).

Site-directed mutagenesis of the 3′ part of *pol* in Moloney murine leukemia virus (Mo-MLV) has resulted in mutant viruses that are incapable of integrating their DNA (8, 16, 32). These mutants can synthesize all three forms of viral unintegrated DNA (full-length linear molecules and circles with one or two LTRs). The replication block in these mutants occurs at some point between viral DNA synthesis and the formation of the intact provirus, presumably at the integration step itself. In previous studies, I have generated two mutations in the IN-coding region of Mo-MLV. The first mutant, MLV-SF1, contains a missense mutation that changes a highly conserved arginine to cysteine at amino acid 114 of IN (8). MLV-SF2 contains a frameshift mutation in IN which also begins at amino acid 114 (16). Viruses derived from both mutations are defective for the formation of normal proviruses, although MLV-SF1 is slightly leaky (16). Schwartzberg et al. (32) have generated a series of deletion mutations in the IN-coding region, and the resulting mutant viruses are also defective for a productive infection. In vitro mutagenesis of the 3′ *pol* region of avian retroviral genomes has generally resulted in viral mutants with a replication-deficient phenotype (18, 29).

One goal of this laboratory is to define the functional domains of the Mo-MLV IN protein. It may be possible to map these domains by the introduction of short oligonucleotides (linkers) into various regions of the IN-coding region (preserving the protein reading frame and deleting little or no original coding sequences). Characterization of the replication phenotype of the resulting viral mutants may be informative in determining the functionally important regions of IN.

By using this approach, I have shown that 10 of 12 in-frame linker insertion mutations introduced into the coding region of IN result in viral mutant pseudotypes capable of synthesizing all forms of viral unintegrated DNA, but incapable of efficiently forming stable proviruses. These data confirm previous studies which indicate that IN plays little or no role in DNA synthesis, but is required for normal provirus formation. In addition, the data suggest that mutations within all but the carboxy terminal 10% of the IN will severely affect the ability of the resulting virus to integrate its DNA. However, two of the linker insertion mutations, which map to the C-terminal region of IN, have no apparent

effect on viral replication and integration. All proviruses generated by the two replication-competent mutants appear to be structurally normal. These results suggest that the carboxy terminus of Mo-MLV IN may not play a functional role in viral DNA integration.

MATERIALS AND METHODS

Cloned DNAs. The plasmid pZAP contains an integrated infectious provirus of Mo-MLV along with flanking cellular rat DNA sequences (20). pMOVpsi− contains an integrated provirus of Mo-MLV which has a deletion in its packaging sequence so that its own genome cannot be packaged into virions (26). pMX262neo was obtained from Mike Scott, University of California, San Francisco, and contains an Mo-MLV-derived retroviral vector missing *pol* and *env* sequences with a *neo* gene derived from Tn*5* fused in-frame with *gag*. pMPE is a pXf3-based plasmid (17) containing the *pol* and *env* genes of Mo-MLV from *Sal*I at nucleotide 3705 to *Cla*I at nucleotide 7674 of the Mo-MLV sequence (34). pUC4-KIXX was purchased from Pharmacia (Uppsala, Sweden) and contains a kanamycin-resistance gene cartridge flanked by polylinkers (2).

Cells and viruses. NIH 3T3 and rat-1 cells were grown in Dulbecco modified Eagle medium containing 10% fetal calf serum. Cells that produced wild-type and mutant MLVs were obtained by cotransfection of the plasmids pZAP and pMOVpsi− or their mutagenized derivatives, along with the vector plasmid pMX262neo, onto NIH 3T3 and rat-1 cells by the transfection protocol of Graham and van der Eb (13) or Chen and Okayama (4). G418-resistant colonies were selected in medium containing 400 to 500 μg of G418 (GIBCO Laboratories, Grand Island, N.Y.) per ml. Mutant and wild-type viruses were harvested from confluent G418-resistant cells and used to infect NIH 3T3 cells in the presence of 8 μg of polybrene (Sigma Chemical Co., St. Louis, Mo.) per ml.

Generation of Mo-MLV IN mutations. The plasmid pMPE, which contains the IN-coding region of Mo-MLV, was the target plasmid for the generation of linker insertion mutations. pMPE was partially cleaved with *Alu*I, *Sma*I, *Pvu*II, *Xba*I, *Bgl*II, and *Kpn*I. pMPE was completely digested with *Sph*I and *Xmn*I (since these enzymes cleave only once in pMPE). Those plasmids cleaved with restriction enzymes which leave overhanging ends were filled in or removed (in the case of 3′ overhanging ends) with T4 DNA polymerase and deoxynucleotide triphosphates to generate blunt ends. The cleaved plasmids were then ligated to *Xho*I linkers of 8, 10, or 12 bp (New England BioLabs, Inc., Beverly, Mass.), depending on the number of base pairs filled in or removed. Those plasmids partially or fully cleaved with a restriction endonuclease which generated blunt ends (e.g., *Alu*I, *Sma*I, *Pvu*II, and *Xmn*I) were ligated to 12-bp linkers so that the integrase reading frame would be maintained. Plasmids cleaved with *Sph*I and *Kpn*I (leaving 3′ overhangs of 4 bp) had 4 bp removed by treatment with T4 DNA polymerase and were thus ligated to 10-bp *Xho*I linkers to maintain the reading frame. Finally, plasmids cut with *Bgl*II and *Xba*I and treated with T4 DNA polymerase (generating a 4-bp duplication) were ligated to 8-bp *Xho*I linkers.

After ligation with *Xho*I linkers, the plasmids were cleaved with *Xho*I and religated to a purified kanamycin resistance gene cassette derived from pUC4-KIXX digested with *Xho*I. These ligations were used to transform *Escherichia coli* C600 in the presence of kanamycin. Kanamycin-resistant colonies were picked, and plasmids were purified by the method of Birnboim and Doly (3) and analyzed by restriction endonuclease mapping for the location of the inserted *Xho*I linker and kanamycin resistance cassette. Plasmids with the desired mutation were cleaved with *Xho*I and religated to remove the kanamycin cassettes. Fragments containing the mutation were purified from agarose gels following digestion with restriction enzymes, and these were ligated into pZAP and pMOVpsi− to generate full-length viral DNAs with a single linker insertion mutation.

Analysis of linker insertion mutations. DNA sequencing to confirm the sequence in the vicinity of each linker insertion mutation was performed using the dideoxy chain termination method (31). pZAP and pMOVpsi− linker insertion mutant DNAs were directly sequenced by denaturation and reannealing with one of four 15- to 17-base oligonucleotide primers (OCS Laboratories, Denton, Tex.) identical to sequences in the Mo-MLV IN-encoding gene. The mutant template primer mixes were extended with a modified T7 polymerase (35) provided in the Sequenase sequencing kit (U.S. Biochemical Corp., Cleveland, Ohio).

Purification of mutant viral DNAs in infected cells was performed by the Hirt fractionation protocol (19). Hirt supernatant DNAs (primarily extrachromosomal) and Hirt pellet DNAs (chromosomal) were purified, cleaved with restriction endonucleases, and subjected to agarose gel electrophoresis. The gels were then alkaline blotted to Zeta-Probe membranes (Bio-Rad Laboratories, Richmond, Calif.) and hybridized to Mo-MLV or *neo* DNA probes by the method of Reed and Mann (30). ^{32}P-labeled probes of high specific activity were generated with an oligolabeling kit (Pharmacia) by the method of Feinberg and Vogelstein (11).

RESULTS

Generation of Mo-MLV IN linker insertion mutations. Twelve linker insertion mutations were constructed spanning the region of Mo-MLV encoding the viral IN (p46^{pol}). Each linker insertion resulted in an insertion of 3 or 4 amino acid codons. The identity of the inserted amino acid codons varied depending on the original cleavage site of the restriction endonuclease and the size of the linker inserted. In *li*5137 and *li*5576, a single amino acid codon was lost from the original sequence as a result of the T4 DNA polymerase removal of four nucleotides in the construction of these mutants. All mutants maintained the original *pol* reading frame, which was confirmed by DNA sequencing in the vicinity of each mutation. The location of each of the linker insertion mutations is shown graphically in Fig. 1. Table 1 indicates how each of the mutations was constructed.

Ability of the mutants to form stable proviruses. Each of the mutated viral DNAs (derived from either pZAP or pMOVpsi−) or wild-type Mo-MLV DNA were cotransfected onto NIH 3T3 cells or rat-1 cells along with the *neo* retroviral vector DNA, pMX262neo. G418-resistant colonies from each mutant and wild-type cotransfected cell line were pooled and amplified. Virus-containing medium was harvested from each line and used to infect NIH 3T3 cells. If the cotransfected helper DNA was derived from pZAP, the virus in the medium was a mixed population of helper virus and *neo* retroviral vector. Otherwise, cells cotransfected with pMOVpsi−-derived constructs generated virus particles with only pseudotyped *neo* vectors. The ability of each mutant-pseudotyped *neo* vector to generate G418-resistant colonies in an infected group of cells was compared to the number of colonies produced by the wild-type pseudotyped *neo* vector.

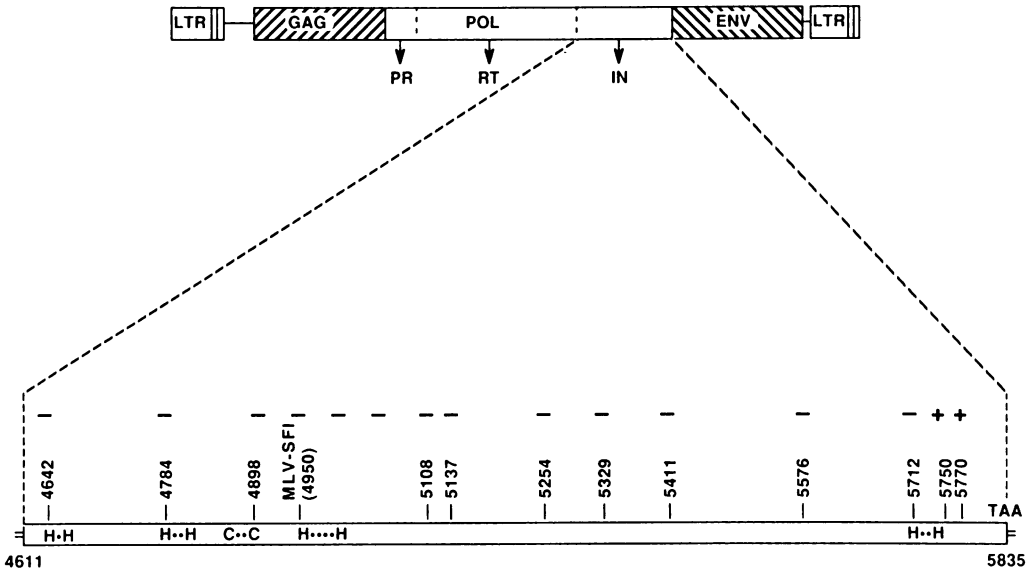

FIG. 1. Location of linker insertion mutations in Mo-MLV. The diagram shows an Mo-MLV provirus with LTRs flanking the three retroviral coding regions. The three mature polypeptides encoded by *pol* are shown below the *pol* coding area. The IN region is expanded in the boxed region below. The nucleotide position of each linker insertion mutation and the MLV-SF1 mutation (8) in the Mo-MLV sequence (34) is indicated above the IN-coding region. The replication phenotype of mutant viruses derived from each mutation is shown above each mutation position. Nucleotides 4611 and 5835 represent the putative 5′ and 3′ ends, respectively, of the IN-coding region. The IN termini have not been experimentally determined but are inferred from previous data (6, 21, 23, 36). Within the box are shown the relative positions of the histidine and cysteine codon pairs. The dots between each member of the pair represent an amino acid codon. Such histidine and cysteine pairs are characteristic of finger-binding proteins (10).

The results of these pseudotyping experiments for the linker insertion mutants are summarized in Table 1. *neo* vectors pseudotyped by wild-type Mo-MLV (pZAP) or packaging-deficient Mo-MLV (pMOVpsi−) usually generated from several hundred to 1,000 colonies per plate after infection with 1 ml of virus-containing medium. Medium from cells transfected with only the *neo* retroviral vector generated no colonies after infection of NIH 3T3 cells. Medium from cells cotransfected with *neo* vector DNA and 10 of 12 linker insertion mutants (derived from both pZAP and pMOVpsi−) usually resulted in no colonies after infection of NIH 3T3 cells (infrequently, one or two colonies could be observed on a plate). However, virus-containing medium derived from two mutants, *li*5750 and *li*5770, could form large numbers of G418-resistant colonies on NIH 3T3 cells, a finding similar to the results seen with wild-type Mo-MLV pseudotyped *neo* vectors. Thus, only 2 of 12 Mo-MLV linker insertion mutants were apparently nondefective for replication by this particular assay.

Ability of the mutants to synthesize viral unintegrated DNA. Previous experiments have shown that viral DNA synthesis is relatively unimpaired in mutants defective in IN function

TABLE 1. Structure and properties of Mo-MLV IN mutants

Mutant[a]	Structure				Properties	
	Parent virus[b]	Insertion site	*Xho*I linker (bp)	Site modification	Formation of G418^r proviruses	Synthesis of unintegrated DNA
*li*4642	M	*Xmn*I	12		−	ND[c]
*li*4784	Z/M	*Pvu*II	12		−	+
*li*4898	Z/M	*Alu*I	12		−	+
MLV-SF1 (4950)	Z	*Sst*II			−	+
MLV-SF2 (4951)	Z	*Sst*II		T4 pol (−2 bp)	−	+
*li*5108	Z/M	*Alu*I	12		−	+
*li*5137	M	*Sph*I	10	T4 pol (−4 bp)	−	ND
*li*5254	Z/M	*Alu*I	12		−	+
*li*5329	Z/M	*Xba*I	8	T4 pol (+4 bp)	−	+
*li*5411	M	*Bgl*II	8	T4 pol (+4 bp)	−	ND
*li*5576	Z/M	*Kpn*I	10	T4 pol (−4 bp)	−	+
*li*5712	Z/M	*Alu*I	12		−	+
*li*5750	Z/M	*Sma*I	12		+	+
*li*5770	M	*Xba*I	8	T4 pol (+4 bp)	+	+

[a] The number indicates the nucleotide of the Mo-MLV sequence (34) after which the linker is inserted. MLV-SF1 contains a C-to-T transition at base 4950 of Mo-MLV that changes a highly conserved Arg codon to a Cys codon (8). MLV-SF2 is a frameshift mutant missing bases 4951 and 4952 (16).

[b] Pseudotyping parent virus. Mutations were placed in pZAP (Z), in pMOVpsi− (M), or in both (Z/M).

[c] ND, Not done.

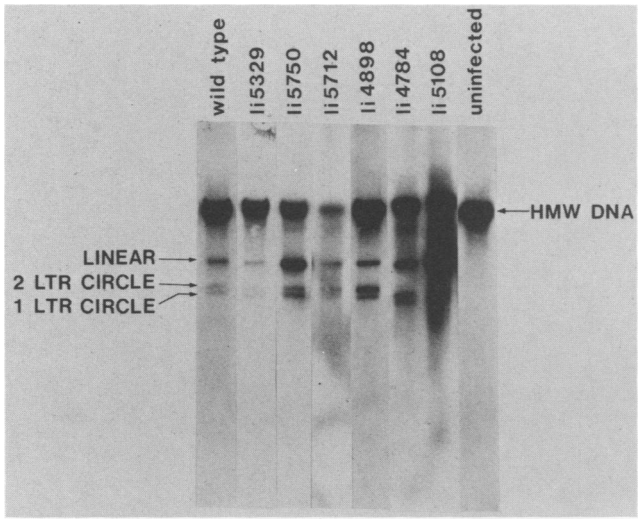

FIG. 2. All linker insertion mutants can synthesize the three forms of viral unintegrated DNA in infected cells. Extrachromosomal DNA was purified from NIH 3T3 cells 24 h after infection with mutant-virus-containing medium. The purified DNAs were electrophoresed in a 0.8% agarose gel, transferred to nylon, and hybridized to a ^{32}P-labeled probe derived from cloned Mo-MLV DNA. The positions of each of the three forms of unintegrated viral DNA are indicated at the left. High-molecular-weight (HMW) chromosomal DNA from infected and uninfected DNA samples hybridizes to the probe because NIH 3T3 cells contain large numbers of endogenous MLV-related proviruses in their genomes.

(8, 32). To confirm this observation for the linker insertion mutants, NIH 3T3 cells were infected with representative mutant viruses and assayed for the ability to synthesize all three forms of viral unintegrated DNA. As summarized in Table 1 and shown in Figure 2, every mutant tested synthesized linear and both circular forms (one and two LTRs) of viral DNA within 24 h after initial infection, a finding similar to that observed in a wild-type Mo-MLV infection. The variation in intensity of viral bands in Fig. 2 is probably a function of the infecting viral titer, which fluctuates extensively in the cotransfected cells. On the basis of the above results, I conclude that viral DNA synthesis is not qualitatively impaired by mutations in the IN-encoding region. (However, I cannot rule out slight quantitative impairments in RT efficiency, since the amounts of infecting virions for each mutant were not directly measured.) Thus, the block in replication of the defective linker insertion mutants is likely to be at some point between viral DNA synthesis and formation of the integrated viral DNA.

Confirmation of replication-competent phenotype of mutants *li*5750 and *li*5770. As noted above, two of the linker insertion mutants were not replication defective. One possible explanation for such a phenotype is that the transfected linker insertion mutant genomes have recombined with endogenous retroviral sequences to generate a wild-type recombinant virus. Schwartzberg et al. (33) have shown that mutant MLV DNAs transfected into murine cells can recombine with endogenous MLV-related genomes by a mechanism called patch repair. I have also occasionally noted this phenomenon when linker insertion mutant DNAs were transfected into NIH 3T3 cells. Consequently, all of the linker insertion mutant DNAs have also been transfected into rat-1 cells, in which recombination should be much less frequent (due to the absence of endogenous MLVs in rat genomes). In fact, I have yet to observe a reversion to wild

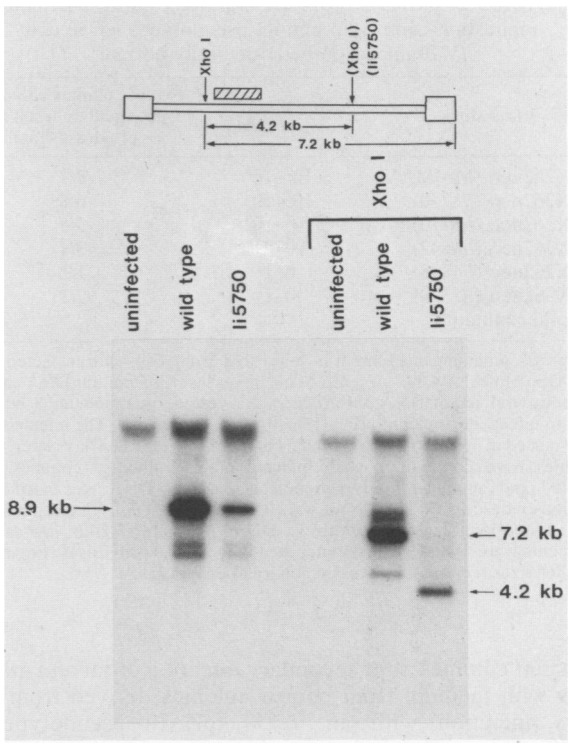

FIG. 3. Replication-competent linker insertion mutant *li*5750 has retained the original mutation in its viral genome. Extrachromosomal DNA was prepared from cells infected with wild-type Mo-MLV and with the replication-competent linker insertion mutant *li*5750. All DNAs were analyzed as described in the legend to Fig. 2, except that some purified DNAs were cleaved with *Xho*I (three lanes at right) before loading on the agarose gel. The diagram at the top shows the expected fragment pattern for *Xho*I-cleaved wild-type Mo-MLV and *li*5750 linear unintegrated DNA. The hatched box indicates the origin of the cloned Mo-MLV DNA fragment (nucleotides 1906 to 3229) used as a labeled probe in these experiments. The uncleaved linear Mo-MLV DNA is 8.9 kb long.

type by a replication-defective linker insertion mutant DNA after transfection into rat cells. In several independent cotransfections into rat and mouse cells, the two mutants *li*5750 and *li*5770 have consistently exhibited a replication-competent phenotype.

Nevertheless, the replication-competent linker insertion mutant *li*5750 was tested by restriction endonuclease mapping of newly synthesized DNA in infected cells. Wild-type Mo-MLV (and patch-repaired mutants) should contain a single *Xho*I site at position 1560. Mutant *li*5750 contains an additional *Xho*I site approximately 4.2 kilobases (kb) 3′ to position 1560. Figure 3 shows that *Xho*I-cleaved viral unintegrated DNA in infected NIH 3T3 cells results in the expected fragment pattern after hybridization to a labeled Mo-MLV fragment extending from nucleotide 1906 to nucleotide 3229. Wild-type viral DNA exhibits a characteristic 7.2-kb band after cleavage with *Xho*I, whereas *li*5750 DNA shows the expected 4.2-kb band after *Xho*I digestion, indicating that the mutant has maintained its linker insertion and not undergone detectable recombination. Some of the same *li*5750-containing medium used to generate these results simultaneously exhibited a replication-competent phenotype in the G418-resistant colony assay. Furthermore, *li*5750 had full viability when assayed for the ability to form a productive infection. This was determined by production of G418-

TABLE 2. pMX262 proviruses formed by wild-type and replication-competent mutant pseudotypes *li*5750 and *li*5770 appear to be structurally normal[a]

Virus stock	No. of G418^r colonies/plate	No. of colonies containing potentially normal *neo* proviruses/total
pMX262neo (Mo-MLV)	10^2–10^3	7/7
pMX262neo (*li*5750)	10^2–10^3	6/6
pMX262neo (*li*5770)	10^2–10^3	5/5
pMX262neo (*li*4642)	0–2	3/4
pMX262neo (*li*4898)	0–1	1/2
pMX262neo (*li*5329)	0–1	2/2
All other mutants	0	

[a] Virus-containing medium was harvested from cells cotransfected with wild-type Mo-MLV DNA or a Mo-MLV linker insertion mutant DNA and the *neo* retroviral vector DNA pMX262neo. Harvested virus medium (1 ml) was used to infect subconfluent NIH 3T3 cells on a 60-mm dish. The infected cells were placed in 500 μg of G418 per ml 24 h after infection. Ten days later, G418 colonies were counted on each infected plate. Individual colonies were isolated and amplified, and high-molecular-weight DNA was purified as previously described. The DNAs were cleaved with *Xba*I, electrophoresed, blotted, and hybridized to *neo* probe. DNAs producing 3.2-kb, *neo*-specific fragments were assumed to contain potentially normal proviruses. Fragments of other sizes were assumed to be abnormal proviruses.

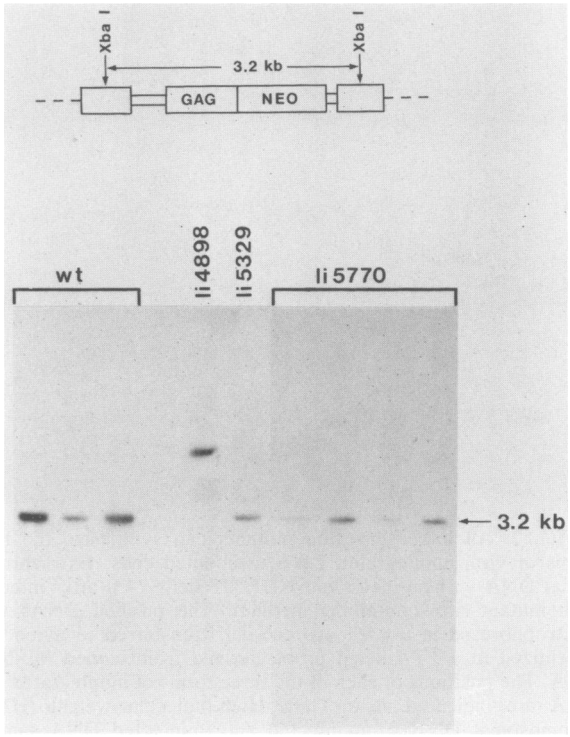

FIG. 4. Replication-competent linker insertion mutants generate structurally normal pseudotyped proviruses while the few pseudotyped proviruses generated by replication-defective mutants are sometimes abnormal. Individual G418-resistant colonies derived from infection with wild-type and mutant-pseudotyped retroviral vector pMX262neo were amplified, and high-molecular-weight DNA was purified. These DNAs were digested with *Xba*I, electrophoresed on a 0.8% agarose gel, blotted to nylon, and hybridized to a 1.1-kb, ^{32}P-labeled *neo* fragment. A schematic map of normally integrated pMX262neo is shown at the top. Digestion of a normal pMX262neo provirus will yield a 3.2-kb *neo*-specific band after cleavage with *Xba*I.

resistant colonies after secondary infection of rat and mouse cells with medium from pooled colonies derived from primary infection with an *li*5750 (pZAP) pseudotype of pMX262neo. The other replication-competent mutant, *li*5770, was constructed in a pMOVpsi− background and therefore could not be tested by this restriction mapping assay (since the helper viral genome would not be present in newly infected cells), but the likelihood of recombination of this mutant genome in each of several independent cotransfections into rat-1 cells is exceedingly remote.

Analysis of proviruses formed by the linker insertion mutants. The formation of G418-resistant colonies by wild-type and mutant-pseudotyped *neo* retroviral vectors is dependent on stable integration of the vector provirus. As indicated in Table 1, colony formation by 10 of the 12 linker insertion mutants does not occur or occurs at a very low frequency. However, the occasional colony that is generated by these replication-defective viruses can be amplified, and the structure of its vector provirus can be analyzed by restriction mapping techniques. The structure of proviruses generated by the two replication-competent mutants can also be analyzed. Past experiments have shown that replication-defective integrase mutants generate few proviruses and that the majority of those that are formed are structurally aberrant by restriction mapping (8, 16).

High-molecular-weight DNA was purified from a number of G418-resistant colonies formed by the replication-competent mutant pseudotypes *li*5750 and *li*5770. A few colonies derived from replication-defective pseudotypes were also obtained, and their high-molecular-weight DNA was purified. The DNAs were cleaved with *Xba*I, which only cleaves once within each LTR of the retroviral vector pMX262neo to generate a 3.2-kb fragment if the vector provirus is integrated in a normal fashion. The observation of a fragment size other than 3.2 kb is evidence that the vector provirus is integrated abnormally. Each of the high-molecular-weight DNAs derived from individual G418-resistant colonies was cleaved with *Xba*I, subjected to agarose gel electrophoresis, blotted to nylon, and hybridized to a labeled *neo* DNA probe. The results are summarized in Table 2 and shown in Fig. 4. All of the proviruses pseudotyped by wild-type Mo-MLV and the two replication-competent linker insertion mutants appeared to be structurally normal by *Xba*I restriction mapping. It was impossible to assess the structure of proviruses generated by seven of the replication-defective mutants because of their inability to form G418-resistant colonies. However, three mutants did produce a few isolated colonies. Two of these mutants generated one clearly abnormal provirus. The third generated two apparently normal proviruses. The results confirm that the replication-competent mutants are integrating their viral DNA in a normal manner. The failure of some of the replication-deficient mutants to consistently form normal proviruses correlates with their inability to generate significant numbers of G418-resistant colonies.

DISCUSSION

The results of the experiments described here are an extension of previous studies (8, 16) and support the argument that the 3′ part of *pol* encodes a protein (p46^{pol} or IN) which is absolutely required for normal retroviral DNA integration. None of the tested mutants were significantly deficient in the formation of any unintegrated forms of viral DNA, confirming that IN has little or no role in viral DNA synthesis. It was hoped that small in-frame linker insertion

mutations spanning the IN-coding region would allow more precise delineation of functionally important IN domains. Previous site-directed mutagenesis studies of this region of *pol* performed by myself and others have not systematically examined the entire Mo-MLV IN-coding region and carefully analyzed the resulting virus replication phenotype (8, 16, 32). The results of this attempt to provide a more comprehensive battery of IN mutants suggested that sequences throughout the protein, with the possible exception of the carboxy terminal 10%, are functionally involved in integration.

Ten of the twelve linker insertion mutants appeared to be defective for integration. Only the two mutations located closest to the carboxy terminus of the IN (*li*5750 and *li*5770) resulted in replication-competent mutant viruses. Lobel and Goff (25) have previously shown that a 12-bp *Eco*RI linker insertion at position 5750 generates a mutant with wild-type replication activity. These results are somewhat analogous to that reported by Katz and Skalka (22), who engineered two mutations into the terminal 37-amino-acid codons of *pol* in avian sarcoma-leukosis virus and found that the resulting mutant viruses were not defective for replication. However, these 37 carboxy-terminal amino acids of *pol* in avian sarcoma-leukosis virus are removed during posttranslational processing and are not present in the mature pp32 endonuclease or the beta subunit of reverse transcriptase (1). There is no evidence that C-terminal processing of Mo-MLV *pol* precursors occurs (21, 36), and it is likely that *li*5750 and *li*5770 contain the inserted amino acids in their mature IN. The *li*5750 and *li*5770 mutations are located immediately 5′ of the overlapping region between the 3′ codons of *pol* and the amino terminal codons of the *env* leader segment. This *pol-env* overlapping region present in many retroviruses appears to undergo rapid sequence change and has no known function (22). Therefore, this region and adjacent domains may be capable of significant alteration without impairment of the viral replication function.

One interesting feature of the IN-coding sequence is the presence of sequence elements characteristic of metal-binding domains found in proteins from a variety of eucaryotic organisms associated with nucleic acid binding (10). The zinc-binding finger proteins have characteristic amino acid sequence motifs which are present to a limited extent in the amino acid sequence of IN. The primary characteristic of the finger proteins is the presence of closely linked cysteine and histidine residue pairs (C2 and H2) which interact to coordinate with a zinc atom. The IN protein has four histidine pairs and 1 cysteine pair (Fig. 1). However, the relative spacing of these pairs does not conform to the finger protein consensus sequence which specifies 12 to 14 amino acid residues between adjacent cysteine and histidine pairs (10). Moreover, only the cysteine pair and one of the histidine pairs appears to be conserved among all retroviruses (38). Since the IN protein does have demonstrated DNA-binding activity (14, 27), it may be of interest to determine whether mutations within these C2 and H2 segments affect integration. Fortuitously, linker insertion mutations have been introduced either within or immediately adjacent to all but one of the C2 and H2 pairs. All of these mutations result in replication-deficient viruses.

The replication-competent mutants and the replication-defective mutants differ quantitatively and qualitatively in provirus formation. The replication-competent mutants *li*5750 and *li*5770 generate large numbers of proviruses, all of which appear to be normal in structure. In contrast, the other mutants produce at least 10^2 to 10^3 fewer proviruses than do wild-type viruses (and replication-competent mutants). Many mutants generated no G418-resistant colonies in any of the experiments, making it impossible to assess proviral structure in these cases. A few mutants generated a small number of colonies (there is probably no difference between these mutants and those mutants that did not generate colonies). Two out of three of these colony-producing mutants generated at least one aberrant provirus, indicating both qualitative deficiencies and quantitative deficiencies in the ability to form proviruses. The apparently normal proviruses generated by replication-deficient linker insertion mutants may be due to leakiness of the mutation, as previously observed for the missense IN mutant, MLV-SF1 (16), or it may be due to the formation of head-to-tail concatemers of viral DNA before recombination with the genomic DNA, as observed with some proviruses generated by the frameshift IN mutant, MLV-SF2 (16). When these two possibilities were investigated by further mapping of the apparently normal proviruses, only one of the six (*li*4642) showed a pattern indicating a concatemer structure (data not shown). Thus, it is possible that most of the mutant proviruses are correctly integrated, although other interpretations are plausible.

Collectively, this battery of linker insertion mutants provides a means for further genetic and biochemical studies of the Mo-MLV integrase. Expression, purification, and biochemical characterization of mutant and wild-type IN will help define regions of the protein that are important for various biochemical and enzymatic activities. Attempts to test for intramolecular complementation between different mutants introduced into the same cell may determine whether the IN has multiple functional domains.

ACKNOWLEDGMENTS

I thank Molly Pettit for skilled technical assistance; Mike Scott and Steve Chang for reagents; Bob Bohannon for help with sequencing; and Joyce Evans for help with the manuscript. I also thank David Steffen for a critical reading of the manuscript.

This work was supported by Public Health Service grant CA41476 from the National Cancer Institute.

LITERATURE CITED

1. **Alexander, F., J. Leis, D. A. Soltis, R. M. Crowl, W. Danho, M. S. Poonian, Y.-C. E. Pan, and A. M. Skalka.** 1987. Proteolytic processing of avian sarcoma and leukosis viruses *pol-endo* recombinant proteins reveals another *pol* gene domain. J. Virol. **61**:534–542.
2. **Barany, F.** 1985. Single-stranded hexameric linkers: a system for in-phase insertion mutagenesis and protein engineering. Gene **37**:111–123.
3. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
4. **Chen, C., and H. Okayama.** 1987. High-efficiency transformation of mammalian cells by plasmid DNA. Mol. Cell. Biol. **7**: 2745–2752.
5. **Cobrinik, D., R. Katz, R. Terry, A. M. Skalka, and J. Leis.** 1987. Avian sarcoma and leukosis virus *pol*-endonuclease recognition of the tandem long terminal repeat junction: minimum site required for cleavage is also required for viral growth. J. Virol. **61**:1999–2008.
6. **Copeland, T. D., G. F. Gerard, C. W. Hixson, and S. Oroszlan.** 1985. Amino- and carboxyl-terminal sequence of Moloney murine leukemia virus reverse transcriptase. Virology **143**:676–679.
7. **Crawford, S., and S. P. Goff.** 1985. A deletion mutation in the 5′ part of the *pol* gene of Moloney murine leukemia virus blocks proteolytic processing of the *gag* and *pol* polyproteins. J. Virol. **53**:899–907.
8. **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.

9. **Duyk, G., M. Longiaru, D. Cobrinik, R. Kowal, P. deHaseth, A. M. Skalka, and J. Leis.** 1985. Circles with two tandem long terminal repeats are specifically cleaved by *pol* gene-associated endonuclease from avian sarcoma and leukosis viruses: nucleotide sequences required for site-specific cleavage. J. Virol. **56**: 589–599.
10. **Evans, R. M., and S. M. Hollenberg.** 1988. Zinc fingers: gilt by association. Cell **52**:1–3.
11. **Feinberg, A., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**:6–13.
12. **Gerard, G. F., and D. P. Grandgenett.** 1975. Purification and characterization of the DNA polymerase and RNase H activities in Moloney murine sarcoma-leukemia virus. J. Virol. **15**:785–797.
13. **Graham, F. L., and A. J. van der Eb.** 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology **52**:456–467.
14. **Grandgenett, D. P., A. C. Vora, and R. D. Schiff.** 1978. A 32,000-dalton nucleic acid-binding protein from avian retrovirus cores possesses DNA endonuclease activity. Virology **89**:119–132.
15. **Grandgenett, D. P., A. C. Vora, R. Swanstrom, and J. C. Olsen.** 1986. Nuclease mechanism of the avian retrovirus pp32 endonuclease. J. Virol. **58**:970–974.
16. **Hagino-Yamagishi, K., L. A. Donehower, and H. E. Varmus.** 1987. Retroviral DNA integrated during infection by an integration-deficient mutant of murine leukemia virus is oligomeric. J. Virol. **61**:1964–1971.
17. **Hanahan, D.** 1983. Studies on transformation of *E. coli* with plasmids. J. Mol. Biol. **166**:557–580.
18. **Hippenmeyer, P. J., and D. P. Grandgenett.** 1985. Mutants of the Rous sarcoma virus reverse transcriptase gene are nondefective in early replication events. J. Biol. Chem. **260**:8250–8256.
19. **Hirt, B.** 1967. Selective extraction of polyoma DNA from infected mouse cell culture. J. Mol. Biol. **26**:365–369.
20. **Hoffman, J. W., D. Steffen, J. Gusella, C. Tabin, S. Bird, D. Cowing, and R. A. Weinberg.** 1982. DNA methylation affecting the expression of murine leukemia viruses. J. Virol. **44**:144–157.
21. **Hu, S. C., D. L. Court, M. Zweig, and J. G. Levin.** 1986. Murine leukemia virus pol gene products: analysis with antisera generated against reverse transcriptase and endonuclease fusion proteins expressed in *Escherichia coli*. J. Virol. **60**:267–274.
22. **Katz, R. A., and A. M. Skalka.** 1988. A C-terminal domain in the avian sarcoma-leukosis virus pol gene product is not essential for viral replication. J. Virol. **62**:528–533.
23. **Kopchick, J. J., J. Harless, B. S. Geisser, R. Killam, R. R. Hewitt, and R. B. Arlinghaus.** 1981. Endodeoxyribonuclease activity associated with Rauscher murine leukemia virus. J. Virol. **37**:274–283.
24. **Leis, J., D. Baltimore, J. M. Bishop, J. Coffin, E. Fleissner, S. P. Goff, S. Oroszlan, H. Robinson, A. M. Skalka, H. M. Temin, and V. Vogt.** 1988. Standardized and simplified nomenclature for proteins common to all retroviruses. J. Virol. **62**:1808–1809.
25. **Lobel, L. I., and S. P. Goff.** 1984. Construction of mutants of Moloney murine leukemia virus by suppressor-linker insertional mutagenesis: positions of viable insertion mutations. Proc. Natl. Acad. Sci. USA **81**:4149–4153.
26. **Mann, R. C., R. C. Mulligan, and D. Baltimore.** 1983. Construction of a retrovirus packaging mutant and its use to produce helper-free defective retrovirus. Cell **33**:153–159.
27. **Misra, T. K., D. P. Grandgenett, and J. T. Parsons.** 1982. Avian retrovirus pp32 DNA-binding protein. I. Recognition of specific sequences on retrovirus DNA terminal repeats. J. Virol. **44**:330–343.
28. **Nissen-Meyer, J., and I. F. Nes.** 1980. Purification and properties of DNA endonuclease associated with Friend leukemia virus. Nucleic Acids Res. **8**:5043–5055.
29. **Panganiban, A. T., and H. M. Temin.** 1984. The retrovirus *pol* gene encodes a product required for DNA integration: identification of a retrovirus int locus. Proc. Natl. Acad. Sci. USA **81**: 7885–7889.
30. **Reed, K. C., and D. A. Mann.** 1985. Rapid transfer of DNA from agarose gels to nylon membranes. Nucleic Acids Res. **13**:7207–7221.
31. **Sanger, F. S., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
32. **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.
33. **Schwartzberg, P., J. Colicelli, and S. P. Goff.** 1985. Recombination between a defective retrovirus and homologous sequences in host DNA: reversion by patch repair. J. Virol. **53**:719–726.
34. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukemia virus. Nature (London) **293**:543–548.
35. **Tabor, S., and C. C. Richardson.** 1987. DNA sequence analysis with a modified T7 DNA polymerase. Proc. Natl. Acad. Sci. USA **84**:4767–4771.
36. **Tanese, N., and S. P. Goff.** 1988. Domain structure of the Moloney murine leukemia virus reverse transcriptase: mutational analysis and separate expression of the DNA polymerase and RNAse H activities. Proc. Natl. Acad. Sci. USA **85**:1777–1781.
37. **Tanese, N., M. J. Roth, and S. P. Goff.** 1986. Analysis of retroviral *pol* gene products with antisera raised against fusion proteins produced in *Escherichia coli*. J. Virol. **59**:328–340.
38. **Toh, H., R. Kikuno, H. Hayashida, T. Miyata, W. Kugimiya, S. Inouye, S. Yuki, and K. Saigo.** 1985. Close structural resemblance between putative polymerase of a *Drosophila* transposable genetic element 17.6 and *pol* gene product of Moloney murine leukemia virus. EMBO J. **4**:1267–1272.
39. **Verma, I. M.** 1975. Studies on reverse transcriptase of RNA tumor viruses. III. Properties of purified Moloney murine leukemia virus DNA polymerase and associated RNase H. J. Virol. **15**:843–854.
40. **Yoshinaka, T., and R. B. Luftig.** 1977. Properties of a proteolytic factor of murine leukemia virus. Cell **12**:709–719.