JOURNAL OF VIROLOGY, Oct. 1990, p. 4709–4717
0022-538X/90/104709-09$02.00/0

Vol. 64, No. 10

Analysis of Mutations in the Integration Function of Moloney Murine Leukemia Virus: Effects on DNA Binding and Cutting

MONICA J. ROTH,[1] PAMELA SCHWARTZBERG,[2] NAOKO TANESE,[2]† AND STEPHEN P. GOFF[2]*

Department of Biochemistry, University of Medicine and Dentistry of New Jersey/Robert Wood Johnson Medical School, Piscataway, New Jersey 08854,[1] and Departments of Biochemistry and Molecular Biophysics and Microbiology, Columbia University College of Physicians and Surgeons, 630 West 168th Street, New York, New York 10032[2]

Received 5 April 1990/Accepted 20 June 1990

The 3′ terminus of the *pol* gene of Moloney murine leukemia virus encodes the integration (IN) protein, required for the establishment of the integrated provirus. A series of six linker insertion mutations and two single-base substitutions were generated within the region encoding the IN protein. Mutations were initially generated within an *Escherichia coli* plasmid expressing the IN protein, and the resulting variants were assayed for DNA-binding activity. Mutations which altered conserved cysteine residues within a potential DNA finger-binding motif resulted in lower or variable DNA binding, which appeared to be the result of variable protein folding. Upon renaturation, these proteins were able to nonspecifically bind DNA in a manner similar to that of the other mutant IN proteins and the parent. When reconstructed back into full-length virus, seven of the eight mutations were lethal. All mutants produced a stable IN protein in virions and mediated normal conversion of the retroviral RNA to its three DNA forms. Fine-structure analysis of the linear double-stranded viral DNA indicated that all seven lethal alterations within the IN protein blocked the formation of the 3′ recessed termini that normally precedes integration.

The key step in the retroviral life cycle that ensures the persistence of the viral genome in the infected cell and its transmission to daughter cells is the integration of a double-stranded DNA copy of the genome into the DNA of the host cell (for recent reviews, see references 21, 40, and 53). Integration of the viral DNA is an efficient and orderly process: specific sites near the ends of the viral genome are joined to the host DNA, so that every integrated provirus has a similar structure and organization (9, 25). These sites consist of short inverted repeats, often imperfect, present at the termini of the long terminal repeats (LTRs) that flank the viral genes (55). During the integration reaction, two characteristic events occur: the two terminal base pairs of the inverted repeats are lost, and a small number of base pairs from the target sequence are duplicated and ultimately flank the integrated provirus (24, 51, 52). These features are shared by retroviruses, other retroposons, and many other transposable elements (50).

A single viral gene product, the integration protein (IN), has been identified as essential for the establishment of the integrated provirus (11, 41, 44, 49). The IN function is encoded near the 3′ end of the viral *pol* gene; mutations in this region have no effect on reverse transcription but can block integration of the viral DNA and subsequent expression of viral genes. The IN region is initially expressed as part of the large *gag-pol* precursor protein (28, 29). In the mammalian retroviruses this precursor is cleaved during virion assembly to generate a separate IN protein of about 40 to 45 kilodaltons (kDa) (58). This protein has been shown to exhibit at least nonspecific DNA-binding activity (34, 46). In the avian retroviruses, the *gag-pol* precursor is only partially cleaved, and the IN function exists both as a free protein termed pp32 and as part of the beta subunit of reverse transcriptase (14, 48). The purified avian proteins have been shown to exhibit DNA binding (22, 39) and DNA endonuclease activity (18–20, 23, 32, 59), with specificity for the termini of the viral LTR (7, 12, 13, 27). The presence of a zinc finger motif—a cysteine-containing sequence capable of folding around a chelated zinc—has been noted and proposed as important for DNA binding (26). Mutagenesis of the Moloney murine leukemia virus (M-MuLV) IN protein has been previously described, but detailed analysis of the biochemical effects of these mutations has not been carried out (10).

Recent work in the M-MuLV system has provided a more detailed definition of the structures and sequences of the DNA intermediates in the integration reaction. Direct analysis of the unintegrated linear viral DNA has shown that the two ends are a mixture of two structures: full-length blunt ends, and termini with the 3′ ends recessed by two nucleotides (5, 45). Kinetic analysis suggested that the terminal nucleotides on the 3′ ends are cleaved in vivo to form recessed 3′ ends and that the IN function is required for this step. More information has followed from in vitro reactions which recapitulate the integration of viral DNA, either endogenous or exogenous, into added target DNA (3, 4, 15, 16). These studies have shown that the linear form of the viral DNA, not the circular forms, is the direct precursor to the provirus; that linear DNAs with recessed 3′ ends are active precursors; and that these 3′ ends are directly joined to target DNA in the strand transfer step, leaving protruding 5′ ends on the viral DNA.

To help define the role of the IN protein in integration, we generated a series of linker insertion and single-base substitutions throughout the domain of the *pol* gene encoding the M-MuLV IN function. The effects of these mutations on the in vitro DNA-binding activity of the protein and on the formation of the recessed 3′ termini of newly synthesized linear viral DNA were examined. Mutations located in the cysteine region, affecting the proposed zinc finger domain, apparently altered the folding of the protein but did not

* Corresponding author.
† Present address: Department of Biochemistry, University of California, Berkeley, CA 94720.

completely abolish the nonspecific DNA binding activity in vitro. Most of these mutations were lethal to virus replication, blocking early events in infection. All mutations that blocked viral integration were found to block the formation of the recessed termini on the linear DNA formed in vivo.

MATERIALS AND METHODS

Cells, viruses, and infections. NIH 3T3 and Rat2 cells were grown in Dulbecco modified Eagle medium (GIBCO) supplemented with 10% (vol/vol) calf serum (Hyclone). Wild-type virus was derived from clone pT11 (33) or pNCA (8). Mutants of M-MuLV were tested for replication competence by the XC assay (47) and by reverse transcriptase assay of the supernatant medium (17) after transformation of NIH 3T3 cells by the DEAE-dextran method (37) with cloned DNA (0.25 μg per 10^5 cells). *dl*5401 virus was obtained from an existing NIH 3T3 producer line (49). Cloned Rat2 cell lines expressing IN^- mutant genomes were established after either Polybrene- or calcium phosphate-mediated cotransformation (61) with pSV2neo DNA (54). Individual colonies were selected for growth in the presence of the drug G418 (GIBCO; 400 μg/ml) and assayed for release of reverse transcriptase activity (17). Virus used for infection was collected from 10-cm dishes containing 5 ml of medium every 12 h. Infections of Rat2 cells were carried out in the presence of Polybrene (8 μg/ml).

Bacteria, plasmids, and extracts. *Escherichia coli* HB101 was used as the host for propagation of all plasmids. The bacterial IN protein expression plasmid pSCB150-18 was described previously (46). Preparation of lysates and isolation of insoluble protein fractions of *E. coli* were done as described previously (57).

Electroblot transfer and binding assay. Southwestern (DNA-protein) blot analysis was performed as described previously (46). In some experiments, after electroblot transfer, the nitrocellulose filter carrying the proteins was soaked in buffer containing 50 mM Tris hydrochloride (pH 8.3), 50 mM dithiothreitol, 1 mM EDTA, and 8 M guanidine hydrochloride for 1 h to denature the proteins. The filter was then washed for 1 h at room temperature in dilution buffer (50 mM Tris-HCl [pH 7.5], 2 mM EDTA, 2 mM dithiothreitol, 0.5 M NaCl, 0.1% Nonidet P-40, and 10% [wt/vol] glycerol), followed by an additional wash at 4°C for at least 24 h in the same buffer, to permit renaturation.

Mutagenesis of pSCB150-18 plasmid DNA. (i) Bisulfite mutagenesis. Plasmid pSCB150-18 DNA was partially digested with *Bsp*MI in the presence of 1 μg of ethidium bromide per ml (42), and the full-length linear DNAs were isolated after electrophoresis (60). The DNA was treated with 2 M sodium bisulfite for 3 h at 37°C and dialyzed as described before (43). The plasmid DNA was recircularized by ligation and used to transform *E. coli* BD1528 (*ung*). Colonies were screened for loss of either the *Apa*LI or *Hin*dIII sites which overlapped the *Bsp*MI cleavage site in the *pol* gene.

(ii) Insertional mutagenesis. After limited digestion of plasmid pSCB150-18 with *Alu*I, the linear DNA was isolated by agarose gel electrophoresis. Then, 12-base-pair (bp) *Eco*RI linkers (sequence CCGGAATTCCGG; New England BioLabs) were added to the termini with T4 DNA ligase. The DNA was digested with *Eco*RI, treated with T4 DNA ligase to form circles, and used to transform bacteria. Mutants were named by the nucleotide position of the mutation, using the numbering system in which the left edge of the left LTR is nucleotide 1.

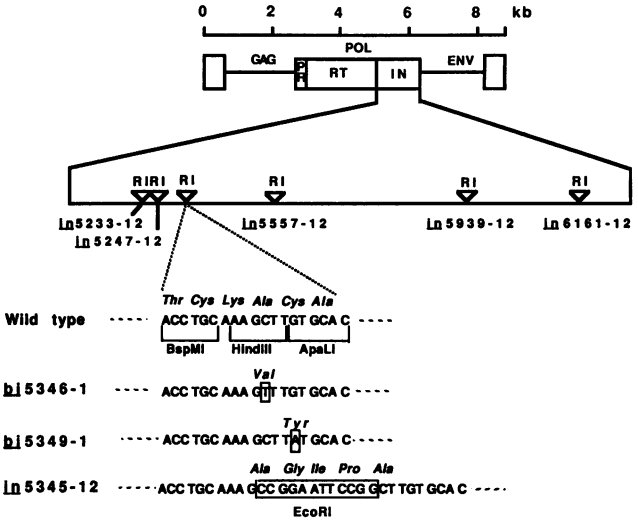

FIG. 1. Positions of mutations affecting the IN protein. M-MuLV proviral DNA is shown at the top; the two boxed regions at both ends of the DNA represent LTR sequences. Enlargement of the *IN* region is shown below. Open triangles above the line indicate positions of 12-bp *Eco*RI (RI) linker insertion mutations. Amino acid and DNA sequences in the cysteine region are shown in the lower half of the figure. RT, Reverse transcriptase; PR, protease.

Antisera and immunoprecipitations. Anti-Pol serum was prepared in rabbits by immunization with the bacterial fusion protein encoded by plasmid pSC1 (58). Goat anti-M-MuLV (serum 81S-107) was obtained from the National Cancer Institute. To detect virion proteins, cultures were labeled with [^{35}S]methionine (Amersham), and the virions were collected, pelleted through a 25% sucrose cushion, and immunoprecipitated as described previously (58). Prestained molecular weight markers were purchased from Sigma Chemical Co.

DNA electroblotting and hybridization. Electroblotting of DNA from sequencing gels and hybridization with oligonucleotide probes were performed as described before (45). The U5 probe D contains the sequence 5′-GAGGAGACT CACTAAC-3′.

RESULTS

Construction of linker insertion mutations in the *IN* domain and analysis of their DNA-binding activity. In previous work, we have expressed the IN protein of M-MuLV in an *E. coli* expression system and demonstrated that both the bacterially derived protein and the natural gene product exhibit DNA-binding activity (46). To identify regions of the IN protein required for binding to DNA, we constructed a set of derivatives of bacterial plasmid pSCB150-18 expressing the IN protein by inserting a 12-bp *Eco*RI linker into *Alu*I sites. Six clones with different inserts in the *IN* region were identified and mapped (Fig. 1). The presence of only a single linker in each mutant was confirmed by detailed restriction mapping, ensuring that each mutation resulted in the insertion of only four amino acids (Table 1).

To test the the ability of the mutant plasmids to encode stable proteins, extracts were prepared after induction of the *trp* promoter, and proteins in the insoluble fraction were analyzed by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (Fig. 2). In each case, the IN protein of

TABLE 1. Summary of mutations in *IN* domain

Virus	Amino acids		Viability[a]	Reverse transcriptase activity[b]
	Wild type	Mutant		
Uninfected Rat 2 cells (control)			−	−
*in*5233-12	QL	QPEFRL	−	+
*in*5247-12	LSF	LSRNSGF	−	++
*in*5345-12	KAC	KAGIPAC	−	++
*in*5557-12	KL	KPEFRL	−	+
*in*5939-12	QAH	QAGIPAH	−	++
*in*6161-12	AAW	AAGIPAW	−	++
*bi*5346-1	A	V	+	+++
*bi*5349-1	C	Y	−	+
Wild type			+	+++

[a] Viability was determined by detection of XC plaques and reverse transcriptase activity after transient transformation of cells with the indicated viral DNAs (see Materials and Methods).

[b] Reverse transcriptase activity present in culture medium from stable producer cell lines expressing the indicated viral DNAs. +++, Wild-type levels; ++, 10 to 25% of wild-type levels; +, 4 to 10% of wild-type levels; −, no detectable activity.

M_r 46,000 could be readily detected by Coomassie blue staining (Fig. 2A); in some cases the mobility of the mutant protein was slightly altered from that of the wild type, suggesting that significant structural changes had been caused by the mutation (e.g., Fig. 2A, lane 6).

Each of the mutant proteins was tested for its ability to bind a mixture of single- and double-stranded DNA by the Southwestern blotting procedure (2, 38). Briefly, proteins were separated by electrophoresis, blotted to nitrocellulose, and challenged with labeled DNA. Most of the mutant proteins retained the parent protein's ability to bind DNA, either with added Mg^{2+} as the divalent cation (Fig. 2B1) or in the presence of EDTA (Fig. 2B2). In general, there was stronger binding with EDTA than with Mg^{2+}. There were significant variations in the amount of DNA bound by different mutant proteins, not corresponding with the amount of protein produced by different clones. The protein from one mutant, *in*5557-12, consistently bound more DNA than those from the other mutants and the wild-type parent (lanes 6). Only one mutant, *in*5345-12, showed a significant defect in binding when normalized to the amount of protein (lanes 5). Similar results were obtained when the binding was performed in higher salt conditions, although here weak binding by mutant *in*5345-12 could be detected (Fig. 2C1 and C2). The mutation in this construct resulted in the insertion of four new amino acids near the N-terminus, increasing the spacing between two cysteine residues from two six amino acids. This region is highly conserved among homologous retroviral *pol* proteins (26) and has the potential to form a zinc finger DNA-binding structure, although not a canonical finger structure (1). These results suggest that this region may be important for binding activity in this assay, affecting either the refolding of the IN protein or its intrinsic binding activity.

Construction of two point mutations near the 5′ end of the *IN* region and further analysis of DNA-binding activity. Because analysis of the insertion mutations suggested that the amino-terminal region of the IN protein was potentially involved in formation of a DNA-binding site, point mutations in this area were also tested. Two mutants were constructed: *bi*5346-1, bearing a highly conservative Ala-to-Val substitution between the two cysteine residues mentioned above, and *bi*5349-1, carrying a change of the second Cys residue to a Tyr (Fig. 1). The proteins encoded by these new mutant plasmids, by the linker insertion plasmids, and by the parent were analyzed for DNA-binding activity after electroblotting (Fig. 2D and E). The IN proteins of the two point mutants bound DNA well, both with (panel D1) and without (panel D2) divalent cation; the Cys residue was therefore not essential for this activity. Experiments with mutant *in*5345-12 gave variable results, depending on the particular preparation of electroblotted filters (Fig. 2). The variation in the binding activity of this mutant could reflect variable refolding of this protein; depending on the experiment, the mutant demonstrated no binding (panels B1 and B2), weak binding (panels C1 and C2), or good binding (panels D1 and D2).

To test whether variations in the renaturation of the proteins were responsible for these results, duplicate gels of the various mutant proteins were prepared and electroblotted as usual. One copy was incubated overnight and then used directly in a binding assay, while another was treated with guanidine hydrochloride and then exposed to renaturation conditions before the assay. Controlled experiments done either directly (Fig. 2E1) or after denaturation and renaturation (Fig. 2E2) showed that the IN proteins with alterations in the cysteines did not refold correctly to recover activity. We conclude that under some conditions, these mutant proteins are capable of binding DNA; however, the ability of the *in*5345-12 protein to renature, compared with that of the proteins from the wild type and the other mutants, is sometimes impaired. When folded correctly, the mutant proteins still exhibit binding activity.

Viability of viral genomes with mutations in the *IN* domain. To test the infectivity of viral genomes carrying site-specific mutations in the *IN* region, fragments carrying each mutation were excised from the expression constructs and used to replace the equivalent fragment of a full-length clone of the viral genome. The resulting cloned DNAs were introduced into NIH 3T3 cells by the DEAE-dextran method (37). Production of replication-competent virus was detected by the XC plaque assay (47) and by measurement of virion-associated reverse transcriptase activity (17). None of the 12-bp insertion mutants gave rise to viable viruses after transfection at 37°C, as judged by the XC plaque assay or the reverse transcriptase assay of the medium (Table 1). Transfection experiments with the two point mutants showed that the *bi*5349-1 mutation, which resulted in the change of the cysteine residue to tyrosine, was similarly lethal. However, mutant *bi*5346-1, with the conservative change of an alanine residue to valine, was viable; the number of XC syncytia induced after DNA transfection was similar to that of the wild-type virus. The level of reverse transcriptase activity detected in the medium was indistinguishable from that in the medium collected from the wild-type producer cells. These results suggested that although several 12-bp insertion mutations did not affect the DNA-binding activity of the fusion protein in vitro, these mutations were nevertheless lethal to the virus in vivo. In addition, the mutation that changed the cysteine residue to tyrosine was lethal in vivo; in the nitrocellulose-binding assay of the IN protein expressed in *E. coli*, however, the cysteine residue did not appear to be critical for DNA binding. Therefore, alterations in the IN protein that have little effect on the DNA-binding activity as assayed in vitro can block its function in vivo. The functions of the IN protein in replication are clearly more demanding than the binding of DNA as measured in vitro.

Construction of cell lines producing IN$^-$ virions. To char-

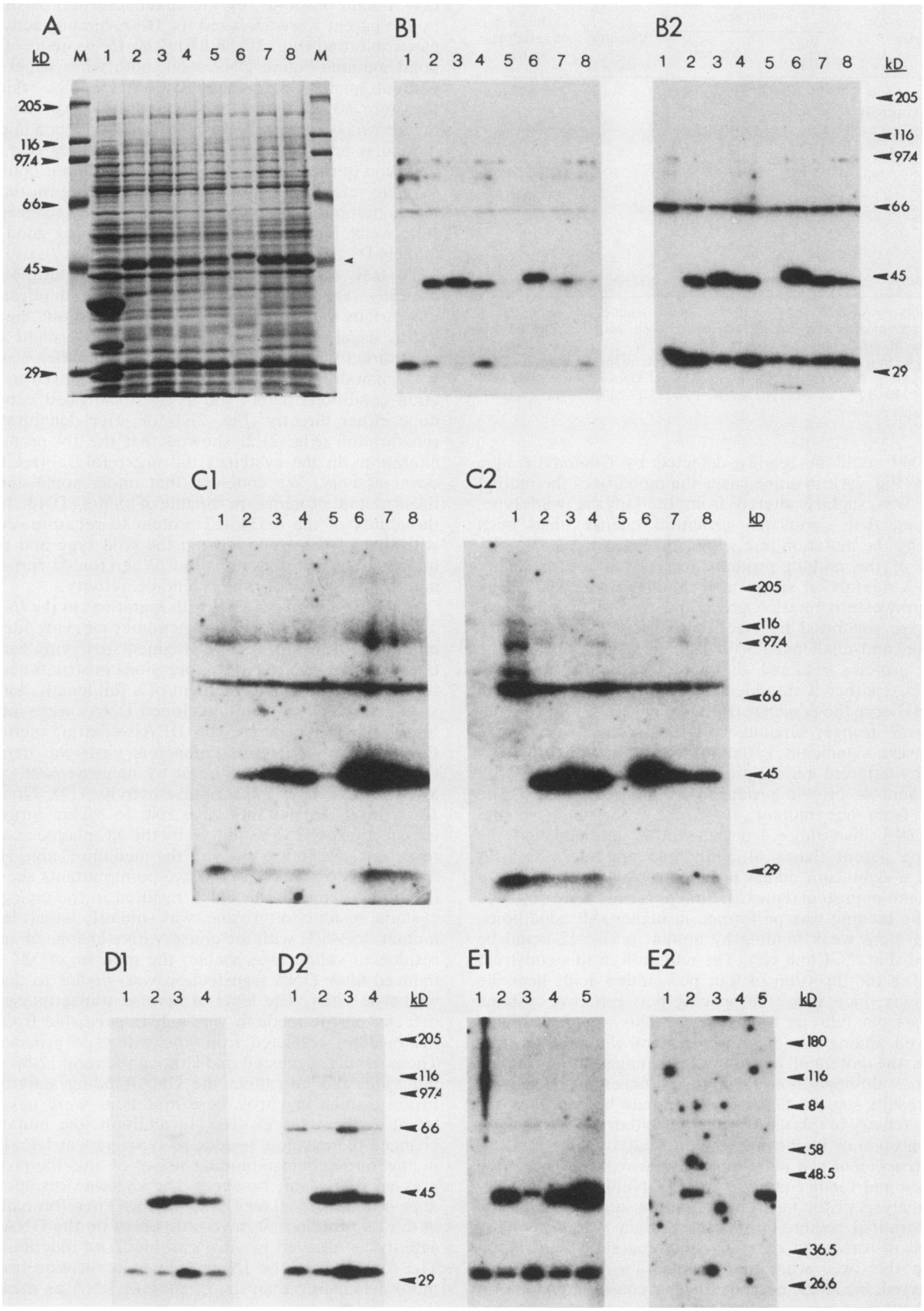
A
kD M 1 2 3 4 5 6 7 8 M
205
116
97.4
66
45
29
B1
1 2 3 4 5 6 7 8
B2
1 2 3 4 5 6 7 8 kD
205
116
97.4
66
45
29
C1
1 2 3 4 5 6 7 8
C2
1 2 3 4 5 6 7 8 kD
205
116
97.4
66
45
29
D1
1 2 3 4
D2
1 2 3 4 kD
205
116
97.4
66
45
29
E1
1 2 3 4 5
E2
1 2 3 4 5 kD
180
116
84
58
48.5
36.5
26.6

FIG. 2. DNA-binding activity of IN proteins with insertion and point mutations. (A to C) Insoluble proteins were prepared from various bacterial lysates; proteins were solubilized in SDS and analyzed by electrophoresis on SDS-polyacrylamide gels. (A) Coomassie stain of the protein gel. In lanes 2 to 8, the 46-kDa (kD) IN protein is indicated by the arrowhead. (B and C) Nitrocellulose-binding assay of the IN proteins. Following electrophoresis, proteins were blotted to nitrocellulose, denatured, renatured, and assayed for DNA-binding activity in the presence of 50 mM NaCl and 5 mM $MgCl_2$ (B1), 50 mM NaCl and 10 mM EDTA (B2), 150 mM NaCl and 5 mM $MgCl_2$ (C1), or 150 mM NaCl and 10 mM EDTA (C2). Samples in panels A to C: lanes 1, pATH3; lanes 2, wild-type plasmid pSCB150-18; lanes 3, *in*5233-12; lanes 4, *in*5247-12; lanes 5, *in*5345-12; lanes 6, *in*5557-12; lanes 7, *in*5939-12; lanes 8, *in*6161-12; lanes M, size markers. Positions of the size standards are indicated. (D) Analysis of proteins from added mutants. Proteins on blots were denatured, renatured, and assayed for DNA-binding activity in the presence of 50 mM NaCl and 5 mM $MgCl_2$ (D1) or 50 mM NaCl and 10 mM EDTA (D2). Lanes 1, pATH3 control cells; lanes 2, pSCB150-18; lanes 3, *in*5345-12; lanes 4, *bi*5349-1. (E) Proteins were electroblotted as before and incubated in dilution buffer alone prior to the DNA-binding assay (E1) or denatured and renatured (E2). Each lane contains insoluble protein pellets from the following plasmids: lane 1, pATH3; lane 2, pSCB150-18; lane 3, *in*5345-12; lane 4, *bi*5349-1; lane 5, *bi*5346-1.

acterize in more detail the block to replication caused by the new mutations in the *IN* domain, we established cell lines expressing the mutant genomes. The complete proviral DNA bearing each mutation was introduced into Rat2 cells by cotransformation with pSV2neo plasmid DNA (54). Individual G418-resistant clones were tested for production of virus by assays of reverse transcriptase activity released into the medium. All the mutant genomes were able to induce release of virion-associated reverse transcriptase activity, suggesting that late stages of the life cycle were normal (Table 1). Virus was collected from these lines and used without dilution to infect fresh NIH 3T3 cells; the number of virions in the preparation, judged from the level of reverse transcriptase activity detected, was approximately 5- to 10-fold less than in a similar preparation of wild-type virus.

Medium was harvested from the infected cells and tested for the appearance of progeny virus by reverse transcriptase assays. Whereas cells infected with wild-type virus became producers of progeny within 1 to 2 days, cells infected with the mutants did not release detectable progeny virus for up to a month postinfection. These results suggested that the mutant virions were unable to establish a productive infection and were blocked in the early phase of the life cycle.

Analysis of the virion proteins from IN$^-$ virus. To examine the viral proteins present in the IN$^-$ mutant virions, producer cells were metabolically labeled with [^{35}S]methionine overnight, the virions were harvested, and the viral proteins were analyzed by SDS gel electrophoresis after immunoprecipitation with an antiserum raised against whole M-MuLV virions (Fig. 3). The abundant *gag* proteins of the virus are the major proteins recognized by this antiserum. In wild-type virions, the initial *gag* gene product, Pr65^{gag}, is normally processed to four smaller cleavage products by the viral protease; in these experiments, high levels of the mature CA protein were detected, showing that, as expected, cleavage of the precursor was essentially complete (Fig. 3, lane 10). In most of the IN$^-$ mutant virions, the predominant *gag* protein detected was also the CA product. The quantity of *gag* protein detected was highly variable from producer line to line; this variation may reflect both the inherent efficiency of assembly of the mutant virion and the level of expression from the particular integrated provirus. In several of the IN$^-$ mutant virions, the Pr65^{gag} precursor protein could still be detected along with various processing intermediates (Fig. 3, lanes 2 to 6 and 9). A reduced rate and final extent of cleavage of the *gag* proteins have been reported previously for several mutants with alterations in the *pol* gene (49). For mutant *in*5557-12, the cleavage was so limited that the levels of the precursor were higher than those of the mature proteins. These results indicate that changes in many portions of the *pol* gene can have some effect on the activity or action of the viral protease. The inability of these mutant viruses to process every *gag* protein in the viral particle, however, is probably not the cause of the lethality of these mutations. These virus were capable of successfully infecting and carrying out reverse transcription after infection of fresh cells (see below).

It is also important to note that all the mutants produced functional levels of *env* gene products. Because of overlap between the *IN* coding region and 5′ portions of the *env* mRNA, the two 3′-proximal mutations constructed here could have affected *env* mRNA expression. In particular, mutation *in*5939-12 is an insertion of 12 bp precisely at the acceptor splice site for the formation of the singly spliced *env* mRNA (31) and might well have disrupted formation of the mRNA and protein. The fact that these two mutants expressed normal amounts of envelope protein and could enter cells and carry out reverse transcription shows that the mutant mRNAs containing the insertions are fully functional.

Analysis of *pol* gene products of the virus. One trivial explanation for the lethality of the IN mutations is that the mutant IN protein might be unstable and not persist in the

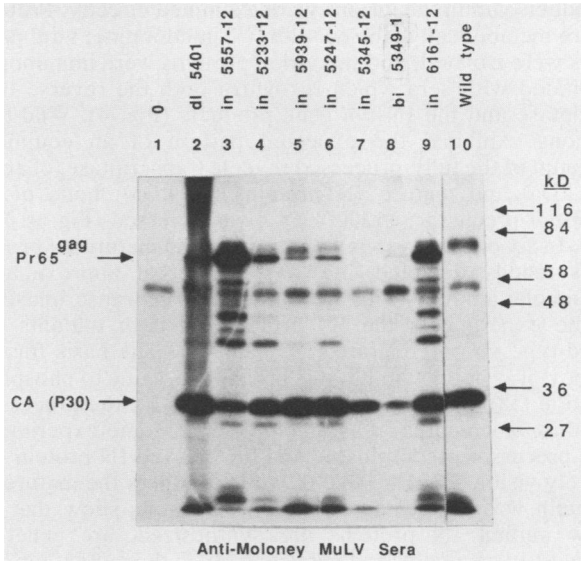

FIG. 3. Analysis of virion proteins. Cells were metabolically labeled with [^{35}S]methionine, and virus was pelleted from the medium. Virion proteins were immunoprecipitated with anti-M-MuLV antiserum and analyzed by SDS-polyacrylamide gel electrophoresis and fluorography. Lane 1, Rat2 cells, uninfected control; lane 2, *dl*5401 in NIH 3T3 cells; lanes 3 to 9, mutants as indicated; lane 10, wild-type M-MuLV. Molecular mass of protein size markers are shown at right (in kilodaltons). Arrows at left indicate the position of the major capsid protein CA and the *gag* polyprotein precursor Pr65^{gag}.

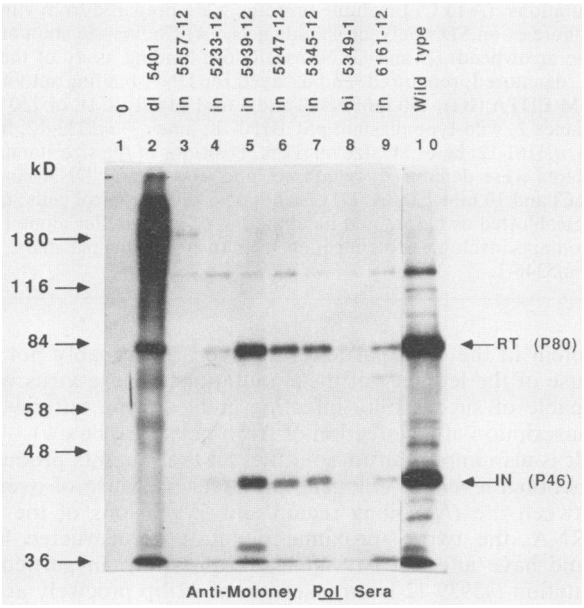

FIG. 4. Analysis of virion *pol* gene products. Cells were metabolically labeled with [^{35}S]methionine, virus was pelleted from the medium, and virion proteins were immunoprecipitated with rabbit anti-pSC1 antiserum and analyzed by SDS-polyacrylamide gel electrophoresis. Lane 1, Rat2 cells, uninfected control; lane 2, *dl*5401 in NIH 3T3 cells; lanes 3 to 9, mutants as indicated; lane 10, wild-type M-MuLV. Positions of protein size markers are shown at left (in kilodaltons). Arrows at right indicate the positions of the reverse transcriptase (RT) and IN proteins.

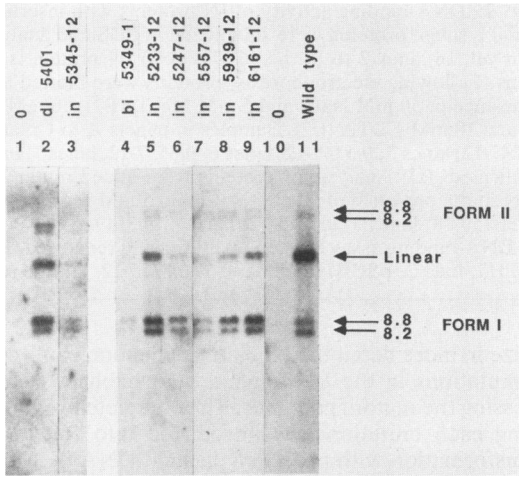

FIG. 5. Southern blot analysis of unintegrated viral DNA. Virus was harvested from producer cell lines and used to infect fresh Rat2 cells. Low-molecular-weight DNA was isolated, separated by electrophoresis on agarose, transferred to nitrocellulose, and probed with M-MuLV sequences. Cells were infected with supernatants from the following lines: lane 1, Rat2 uninfected control; lane 2, *dl*5401 in NIH 3T3 cells; lanes 3 to 9, mutants as indicated; lane 10, mock-infected control; lane 11, wild-type M-MuLV control. Lanes 1 to 3 were a 7-day exposure. Lanes 4 to 11 were a 1-day exposure. Sizes are indicated at the right (in kilobases).

mature virion. To examine this possibility, the *pol* gene products within the virions were examined directly. Proteins were metabolically labeled with [^{35}S]methionine, viral particles were isolated, and the virion proteins were immunoprecipitated with sera which recognize both the reverse transcriptase and the IN *pol* gene products (Fig. 4). Wild-type virions exhibited the expected pattern of an equimolar amount of the fully-processed reverse transcriptase (of about 80 kDa) and mature IN proteins (46 kDa); none of the *gag-pol* precursor, Pr200$^{gag\text{-}pol}$, was detected (Fig. 4, lane 10). In six of the seven mutant virions, the mature IN protein was readily detected and was present at approximately equimolar amounts with the mature reverse transcriptase. In some experiments the IN proteins of both mutants and wild-type virions migrated as a doublet. The basis for this behavior is unknown, but it is probably not due to phosphorylation (58). For insertion mutant *in*5557-12, the processing of the *pol* precursor was incomplete; in some experiments the precursor predominated and the cleaved IN protein was barely visible (Fig. 4, lane 3), while in others the mature IN protein was readily detected. These results show that the new mutant IN proteins are synthesized, are generally processed correctly, and persist stably in the mutant virions. As previously described, virions of the deletion mutant *dl*5401 contained no detectable IN protein (58).

Analysis of the viral DNA synthesized after infection by IN$^-$ virus. To determine whether the mutant virions could initiate an infection and mediated the reverse transcription of the viral RNA into DNA, we examined the viral DNA in recently infected cells. Virus preparations of the various IN$^-$ mutants were collected from the stable producer cells and used to infect fresh Rat2 cells. The low-molecular-weight DNA was isolated 30 h postinfection, separated by electrophoresis, and analyzed by Southern blot hybridization (Fig. 5).

In a wild-type M-MuLV infection, three viral DNA products were detected: an 8.8-kb linear DNA, an 8.8-kb circular DNA, and an 8.2-kb circular species (Fig. 5, lane 11). The linear species was the predominant form at this early time in infection; the amount of circular forms increased at later time points (data not shown). Infection with the deletion mutant *dl*5401 generated the same three proviral DNA forms, although the ratio of circular to linear forms was higher than in the wild type (lane 2), as described previously (45, 49). All of the IN$^-$ linker insertion and point mutants were capable of infecting cells and synthesizing the three proviral DNA products (lanes 3 to 9). The ratio of circular to linear DNA species again was generally increased, favoring the circular forms, as for mutant *dl*5401. Similar observations have been made after infection by mutants defective in integration due to changes in the LTR termini (8). For mutant *in*5557-12, the incomplete processing of the *gag* and *pol* gene products apparently had little effect on the ability of the virus to reverse transcribe its RNA genome (Fig. 5, lane 7). Interestingly, although the virions of the insertion mutant *in*5345-12 contained high levels of reverse transcriptase activity as well as reverse transcriptase protein (Fig. 4, lane 7), this mutant consistently synthesized a disproportionately low level of proviral DNA.

Analysis of the viral DNA formed after infection by the various mutants showed that all the mutations were still present. All the viral DNAs containing the linker insertions were sensitive to digestion with *Eco*RI, and the DNAs from the bisulfite mutants were resistant to digestion with *Apa*LI (data not shown). The results show that all the IN$^-$ mutants were genetically stable and at least largely competent to carry out reverse transcription of all the viral DNA forms.

Analysis of the termini of linear viral DNA synthesized after infection by IN$^-$ virus. Recent analyses of unintegrated linear

retroviral DNAs have demonstrated that a mixture of two terminal structures are present: some termini are blunt-ended and others have 3′ ends recessed by two nucleotides (5, 45). The IN$^-$ deletion mutant *dl*5401 was found to be unable to form the recessed ends, suggesting that the IN protein was required for and might directly mediate the cleavage of these strands. To test whether the new mutant IN proteins could mediate this early step in the integration process, we examined the terminal structures of the linear DNAs formed after infection by the IN$^-$ viruses.

The low-molecular-weight DNA from freshly infected cells was first isolated and digested with restriction enzymes to generate small terminal fragments. The resulting fragments were then separated by electrophoresis on a sequencing gel, electroblotted onto nylon filters, and probed with oligonucleotides specific for each of the strands at the two termini of the linear DNA (Fig. 6). DNA from wild-type virus showed the expected mixture of blunt and recessed 3′ ends (lanes 1, 11); deletion mutant *dl*5401 showed only the blunt ends, as seen previously (lane 2). DNA from each of the six linker insertion mutants and the defective point mutant *bi*5349-1 showed only the blunt termini, with no detectable formation of the 3′ recessed ends (lanes 4 to 10). Thus, every IN$^-$ mutant was defective in the cleavage process, the first detectable step leading to the integrated provirus. No region of the protein was singled out as dispensable for this early event but needed for later events; there were no partially functional mutants, as tested at 37°C.

In confirmation of the Southern blot analysis above, the IN$^-$ viruses generally directed the formation of a higher proportion of the LTR-LTR junction fragment than terminal fragments (lanes 4 to 10). The higher abundance of that fragment reflects the higher abundance of circular DNAs over linear seen in the absence of integration.

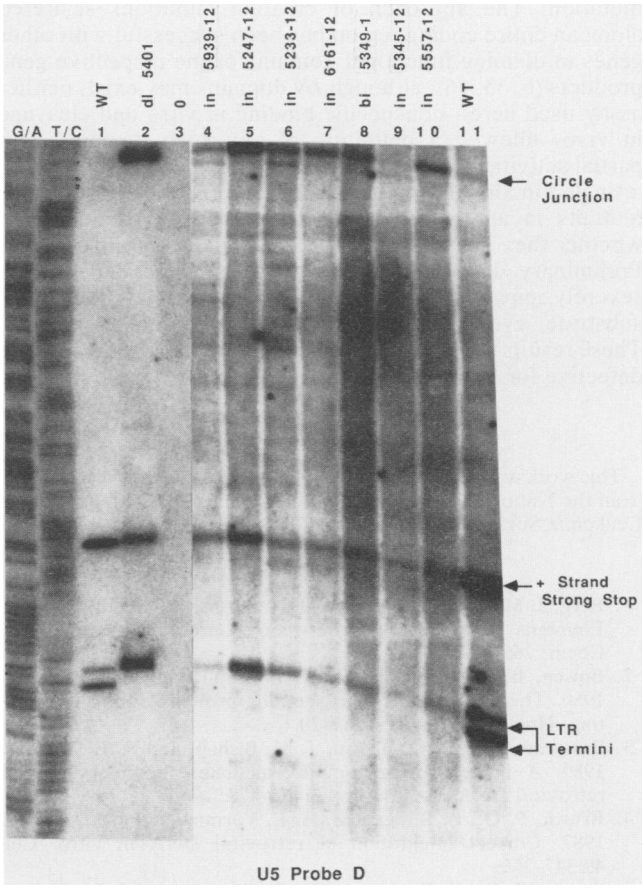

FIG. 6. Analysis of termini from a collection of mutant IN$^-$ viruses. Wild-type M-MuLV and mutant IN$^-$ viruses were collected from producer cells and used to infect Rat2 cells. Cells were lysed 30 h postinfection, and terminal DNA structures were analyzed after blotting as described in Materials and Methods. The filter was probed to detect the 3′ terminus of U5. For IN$^-$ virus, proviral DNA from approximately 6×10^7 to 9×10^7 cells was analyzed; for wild-type virus, DNA from 1×10^7 to 2×10^7 cells was used. Lanes 1 and 11, Wild-type (WT) virus. Lane 2, *dl*5401 virus (IN$^-$). Lane 3, DNA isolated after a mock infection. Lanes 4 to 10, Mutants as indicated. The identities of the various DNA species, based on the size of the fragments, are indicated at the side of the panel. Marker DNAs derived from Maxam and Gilbert sequencing reactions (36) (G+A and T+C lanes) are at the left.

DISCUSSION

In these studies, we have examined the effects of various mutations in the M-MuLV IN protein, both in vitro, with a DNA-binding assay, and in vivo, by measuring the ability of the mutant viruses to proceed through the life cycle. Mutations altering the IN protein had complicated effects on the binding activity. Most of the mutations showed very little effect on the potent, nonspecific DNA-binding activity of the protein. Two mutations, both mapping in the zinc finger region near the N-terminal third of the protein, did reduce the binding in some assays, although two other mutations in this region had no effect. Alterations changing the distance between two cysteines (*in*5345-12) or eliminating one of these two cysteines (*bi*5349-1) often profoundly decreased binding, suggesting that the cysteines may function in forming a DNA-binding structure, but their effects were variable and critically dependent on the way the protein was handled before assay, as is often seen for renatured enzymes (30). Thus, the mutations may act largely by affecting the ability of the protein to renature rather than by directly affecting the inherent binding of the protein. Furthermore, the activity assayed after blotting may require only partial refolding and may not reflect the true sequence specificity of the native protein. Other assays may be required to define functional domains of the protein.

The effects of these mutations as judged by our other assays in vivo were more severe. All of the linker insertion mutations were lethal to the virus, suggesting that the IN protein carries out far more specific functions than the simple DNA binding measured in vitro. Of the bisulfite mutations, only the extremely conservative Ala-to-Val mutation had no effect on virus replication. In all the mutants, the ability of the virus to enter the cell and synthesize the linear viral double-stranded DNA was not altered; integration itself was apparently blocked. Fine-structure analysis of the termini of the viral DNA formed in vivo indicated that the linear molecules from IN$^-$ mutant virus were all blunt-ended. These results support the model (5, 45) that the IN protein is required for the formation of the 3′ recessed termini and that this structure is in turn required for integration of the DNA.

The process of integration seems to be divisible into two steps: cleavage of the termini of the viral DNA, and then strand transfer of the 3′ ends into the target DNA, leaving nicks in the target (5, 16). All of the mutants analyzed here were blocked in the first step, the cleavage of the viral DNA, and we have not been able to separate these two steps by

mutation. The approach of creating mutations scattered along an entire coding region has been successful with other genes in defining functional domains of the respective gene products (6, 35, 56); although *IN* domains may exist, neither assay used here—nonspecific binding in vitro and cleavage in vivo—allowed a definition of any such domains with partial activities. It is possible that the IN protein is folded as a single functional protein. We are currently analyzing these mutants in an in vitro integration reaction to determine whether they are blocked at separate stages of the process. Preliminary data suggest that most of these mutants are severely impaired in their ability to integrate an exogenous substrate, even when the recessed 3′ ends are preformed. These results suggest that the mutants we have generated are defective for both steps of the integration process.

ACKNOWLEDGMENTS

This work was supported by Public Health Service grant CA30488 from the National Cancer Institute. M.J.R is a special fellow of the Leukemia Society of America, Inc.

LITERATURE CITED

1. **Berg, J. M.** 1990. Zinc fingers and other metal-binding domains. Elements for interactions between macromolecules. J. Biol. Chem. **265**:6513–6516.
2. **Bowen, B., J. Steinberg, U. K. Laemmli, and H. Weintraub.** 1980. The detection of DNA-binding proteins by protein blotting. Nucleic Acids Res. **8**:1–20.
3. **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.
4. **Brown, P. O., B. Bowerman, H. E. Varmus, and J. M. Bishop.** 1987. Correct integration of retroviral DNA in vitro. Cell **49**:347–356.
5. **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.
6. **Chen, M., and M. S. Horwitz.** 1989. Dissection of functional domains of adenovirus DNA polymerase by linker-insertion mutagenesis. Proc. Natl. Acad. Sci. USA **86**:6116–6120.
7. **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.
8. **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.
9. **Dhar, R., W. McClements, L. Enquist, and G. Vande Woude.** 1980. Nucleotide sequences of integrated Moloney sarcoma provirus long terminal repeats and their host and viral junctions. Proc. Natl. Acad. Sci. USA **77**:3937–3941.
10. **Donehower, L. A.** 1988. Analysis of mutant Moloney murine leukemia viruses containing linker insertion mutations in the 3′ region of *pol*. J. Virol. **62**:3958–3964.
11. **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.
12. **Duyk, G., J. Leis, M. Longiaru, and A. M. Skalka.** 1983. Selective cleavage in the avian retroviral long terminal repeat sequence by the endonuclease associated with the alpha-beta form of avian reverse transcriptase. Proc. Natl. Acad. Sci. USA **80**:6745–6749.
13. **Duyk, G., M. Longiaru, D. Cobrinik, R. Kowal, P. D. Haseth, 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**:586–599.
14. **Eisenman, R. N., W. S. Mason, and M. Linial.** 1980. Synthesis and processing of polymerase proteins of wild-type and mutant avian retroviruses. J. Virol. **36**:62–78.
15. **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.
16. **Fujiwara, T., and K. Mizuuchi.** 1988. Retroviral DNA integration: structure of an integration intermediate. Cell **54**:497–504.
17. **Goff, S. P., P. Traktman, and D. Baltimore.** 1981. Isolation and properties of Moloney murine leukemia virus mutants: use of a rapid assay for release of virion reverse transcriptase. J. Virol. **38**:239–248.
18. **Golomb, M., and D. P. Grandgenett.** 1979. Endonuclease activity of purified RNA-directed DNA polymerase from avian myeloblastosis virus. J. Biol. Chem. **254**:1606–1613.
19. **Golomb, M., D. P. Grandgenett, and W. Mason.** 1981. Virus-coded DNA endonuclease from avian retrovirus. J. Virol. **38**:548–555.
20. **Grandgenett, D. P., M. Golomb, and A. C. Vora.** 1980. Activation of an Mg^{2+}-dependent DNA endonuclease of avian myeloblastosis virus alpha-beta DNA polymerase by in vitro proteolytic cleavage. J. Virol. **33**:264–271.
21. **Grandgenett, D. P., and S. R. Mumm.** 1990. Unraveling retrovirus integration. Cell **60**:3–4.
22. **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.
23. **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.
24. **Hughes, S. H., A. Mutschler, J. M. Bishop, and H. E. Varmus.** 1981. A Rous sarcoma virus provirus is flanked by short direct repeats of a cellular DNA sequence present in only one copy prior to integration. Proc. Natl. Acad. Sci. USA **78**:4299–4303.
25. **Hughes, S. H., P. R. Shank, D. H. Spector, H.-J. Kung, J. M. Bishop, H. E. Varmus, P. K. Vogt, and M. L. Breitman.** 1978. Proviruses of avian sarcoma virus are terminally redundant, co-extensive with unintegrated linear DNA, and integrated at many sites. Cell **15**:1397–1410.
26. **Johnson, M. S., M. A. McClure, D.-F. Feng, J. Gray, and R. F. Doolittle.** 1986. Computer analysis of retroviral *pol* genes: assignment of enzymatic functions to specific sequences and homologies with nonviral sequences. Proc. Natl. Acad. Sci. USA **83**:7648–7652.
27. **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.
28. **Kopchick, J. J., G. A. Jamjoom, K. F. Watson, and R. B. Arlinghaus.** 1978. Biosynthesis of reverse transcriptase from Rauscher murine leukemia virus by synthesis and cleavage of a *gag-pol* readthrough viral precursor polyprotein. Proc. Natl. Acad. Sci. USA **75**:2016–2020.
29. **Kopchick, J. J., W. L. Karshin, and R. B. Arlinghaus.** 1979. Tryptic peptide analysis of *gag* and *gag-pol* gene products of Rauscher murine leukemia virus. J. Virol. **30**:610–623.
30. **Lacks, S. A., and S. S. Springhorn.** 1980. Renaturation of enzymes after polyacrylamide gel electrophoresis in the presence of sodium dodecylsulphate. J. Biol. Chem. **255**:7467–7471.
31. **Lazo, P. A., V. Prasad, and P. N. Tsichlis.** 1987. Splice acceptor site for the *env* message of Moloney murine leukemia virus. J. Virol. **61**:2038–2041.
32. **Leis, J., G. Duyk, S. Johnson, M. Longiaru, and A. M. Skalka.** 1983. Mechanism of action of the endonuclease associated with the alpha-beta and beta-beta forms of avian RNA tumor virus reverse transcriptase. J. Virol. **45**:727–739.
33. **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.
34. **Luc, K.-C., T. D. Gilmore, and A. T. Panganiban.** 1987. The

spleen necrosis virus *int* gene product expressed in *E. coli* has DNA binding activity and mediates *att* and U5-specific DNA multimer formation *in vitro*. Virology **157**:127–136.

35. **Lyman, S. D., and L. R. Rohrschneider.** 1987. Analysis of functional domains of the v-*fms*-encoded protein of Susan McDonough strain feline sarcoma virus by linker insertion mutagenesis. Mol. Cell. Biol. **7**:3287–3296.
36. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
37. **McCutchan, J. H., and J. S. Pagano.** 1968. Enhancement of the infectivity of simian virus 40 deoxyribonucleic acid with diethylaminoethyl-dextran. J. Natl. Cancer Inst. **41**:351–357.
38. **Miskimins, W. K., M. P. Roberts, A. McClelland, and F. H. Ruddle.** 1985. Use of a protein-blotting procedure and a specific DNA probe to identify nuclear proteins that recognize the promoter region of the transferrin receptor gene. Proc. Natl. Acad. Sci. USA **82**:6741–6744.
39. **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.
40. **Panganiban, A. T.** 1985. Retroviral DNA integration. Cell **42**:5–6.
41. **Panganiban, A. T., and H. M. Temin.** 1984. The retrovirus *pol* gene encodes a product required for DNA integration: identification of a retroviral *int* locus. Proc. Natl. Acad. Sci. USA **81**:7885–7889.
42. **Parker, R. C., R. M. Watson, and J. Vinograd.** 1977. Mapping of cloned circular DNAs by cleavage with restriction endonucleases and calibration by agarose gel electrophoresis. Proc. Natl. Acad. Sci. USA **74**:851–855.
43. **Peden, K. W. C., and D. Nathans.** 1982. Local mutagenesis within deletion loops of DNA heteroduplexes. Proc. Natl. Acad. Sci. USA **79**:7214–7217.
44. **Quinn, T. P., and D. P. Grandgenett.** 1988. Genetic evidence that the avian retrovirus DNA endonuclease domain of *pol* is necessary for viral integration. J. Virol. **62**:2307–2312.
45. **Roth, M. J., P. 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.
46. **Roth, M. J., N. Tanese, and S. P. Goff.** 1988. Gene product of Moloney murine leukemia virus required for proviral integration is a DNA-binding protein. J. Mol. Biol. **203**:131–139.
47. **Rowe, W. P., W. E. Pugh, and J. W. Hartley.** 1970. Plaque assay techniques for murine leukemia viruses. Virology **42**:1136–1139.
48. **Schiff, R. D., and D. P. Grandgenett.** 1978. Virus-encoded origin of a 32,000-dalton protein from avian retrovirus cores: structural relatedness of p32 and the β polypeptide of the avian retrovirus DNA polymerase. J. Virol. **28**:279–291.
49. **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.
50. **Shimotohno, K., S. Mizutani, and H. M. Temin.** 1980. Sequence of retrovirus provirus resembles that of bacterial transposable elements. Nature (London) **285**:550–554.
51. **Shoemaker, C., S. 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.
52. **Shoemaker, C., J. Hoffman, S. P. Goff, and D. Baltimore.** 1981. Intramolecular integration within Moloney murine leukemia virus DNA. J. Virol. **40**:164–172.
53. **Skalka, A. M.** 1989. Integrative recombination of retroviral DNA, p. 701–724. *In* R. S. Kucherlapati and G. R. Smith (ed.), Genetic recombination, American Society for Microbiology, Washington, D.C.
54. **Southern, P. J., and P. Berg.** 1982. Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. J. Mol. Appl. Genet. **1**:327–341.
55. **Swanstrom, R., W. J. DeLorbe, J. M. Bishop, and H. E. Varmus.** 1981. Nucleotide sequence of cloned unintegrated avian sarcoma virus DNA: viral DNA contains direct and inverted repeats similar to those in transposable elements. Proc. Natl. Acad. Sci. USA **78**:124–128.
56. **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.
57. **Tanese, N., M. J. Roth, and S. P. Goff.** 1985. Expression of enzymatically active reverse transcriptase in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **82**:4944–4948.
58. **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.
59. **Terry, R., D. A. Soltis, M. Katzman, D. Cobrinik, J. Leis, and A. M. Skalka.** 1988. Properties of avian sarcoma-leukosis virus pp32-related *pol*-endonucleases produced in *Escherichia coli*. J. Virol. **62**:2358–2365.
60. **Vogelstein, B., and D. Gillespie.** 1979. Preparative and analytical purification of DNA from agarose. Proc. Natl. Acad. Sci. USA **76**:615–619.
61. **Wigler, M., R. Sweet, G. K. Sim, B. Wold, A. Pellicer, E. Lacy, T. Maniatis, S. Silverstein, and R. Axel.** 1979. Transformation of mammalian cells with genes from procaryotes and eucaryotes. Cell **16**:777–785.