Sequence-Specific Binding of DNA by the Moloney Murine Leukemia Virus Integrase Protein

PAUL A. KROGSTAD¹ AND JAMES J. CHAMPOUX^{2*}

Departments of Pediatrics¹ and Microbiology,² University of Washington, Seattle, Washington 98195

Received 14 August 1989/Accepted 2 March 1990

Genetic studies have indicated that integration of retroviral DNA into the host genome depends on the presence of the inverted repeats at the free termini of the long terminal repeats on the unintegrated DNA and on the product of the 3' end of the *pol* gene (the integrase [IN] protein). While the precise function of the Moloney murine leukemia virus IN protein is uncertain, others have shown that it is a DNA-binding protein and functions in the processing of the inverted repeats prior to integration. By using site-directed mutagenesis, we cloned and expressed the IN protein in *Escherichia coli*. Crude extracts of total cellular protein were fractionated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis, transferred to nitrocellulose filters, denatured in guanidine, renatured, and incubated with oligonucleotide probes. Single- and double-stranded oligonucleotides corresponding to the termini of unintegrated linear viral DNA were specifically bound by the IN protein in this assay. These data suggest that the role of the Moloney IN protein in the early steps of integration involves sequence-specific recognition of the DNA sequences found at the ends of the long terminal repeats.

During retroviral replication, the linear double-stranded DNA (dsDNA) product of reverse transcription is integrated into the host cell genome (2, 10, 11, 22). The biochemical mechanism of this illegitimate recombination event is unknown, but it characteristically involves a loss of base pairs from the termini of the long terminal repeats (LTRs) of the viral DNA and the generation of short direct repeats (4 to 6 base pairs) in the host cell DNA at the site of integration (33). Genetic studies have indicated that integration is dependent on the product of the 3' end of the *pol* gene (encoding the integrase [IN] protein) (7, 16, 28); cells infected with certain *pol* mutants of Moloney murine leukemia virus (M-MuLV) contain the usual linear and circular DNA intermediates in the cytoplasm and nucleus, but few, generally aberrant, integrants are observed (6, 7). In recent studies, the DNA intermediates of integration have been isolated and characterized (3, 12, 27). Extracts of cells infected with wild-type M-MuLV are enriched for a linear viral DNA species in which both 3' ends are shortened by two bases. In contrast, viral DNA isolated from cells infected with M-MuLV *pol* mutants defective in the IN domain are predominantly blunt ended and presumably represent full-length reverse transcription products (3, 27). While the precise function of the M-MuLV IN protein is uncertain, others have shown that it has nonspecific DNA binding activity with a preference for single-stranded (ss) over dsDNA (26).

We have cloned and expressed the IN protein in *Escherichia coli* and probed protein blots with synthetic oligonucleotides corresponding to the inverted repeats (IRs) found at the termini of the viral LTRs. The results provide evidence for sequence-specific DNA binding by the M-MuLV IN protein and suggest an additional role for oligonucleotides in the study of DNA-protein interactions.

MATERIALS AND METHODS

General methods and materials. Restriction endonucleases, Klenow fragment, T4 DNA polymerase, and T4

polynucleotide kinase were purchased from Bethesda Research Laboratories, Inc., U.S. Biochemicals Corp., and New England BioLabs, Inc., and were used according to the specifications of the manufacturers. Indoleacrylic acid, isopropyl- β -D-thiogalactopyranoside, HEPES (*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid), and guanidine-HCl were purchased from Sigma Chemical Co. Radioactive nucleotides were purchased from Dupont, NEN Research Products. Plasmid DNA was isolated by standard techniques (1) from transformed *E. coli* DH5 α . Single-stranded DNA templates for site-directed mutagenesis and DNA sequencing were extracted from phage particles produced by *E. coli* strains containing M13-ori plasmids upon infection with the helper phage M13K07. Oligonucleotides were end labeled with [γ -³²P]ATP (3,000 Ci/mmol) and T4 polynucleotide kinase in 100 mM Tris hydrochloride (pH 7.5)–5 mM MgCl₂–1 mM dithiothreitol. The kinase reactions were generally carried out at 0°C for 2 h. Unincorporated nucleotides were removed by performing two successive ethanol precipitations in the presence of 2.5 M ammonium acetate.

Oligonucleotide preparation. Oligonucleotides were synthesized by using a Biosearch Model 8600 synthesizer. They were purified by electrophoresis through 20% polyacrylamide–8 M urea gels containing 89 mM Tris base–89 mM borate–2.5 mM EDTA (TBE). The full-length products were eluted from the gel, ethanol precipitated, and quantified by UV absorbance. Oligonucleotides were prepared corresponding to both the plus (same polarity as the RNA genome) and minus strands of the inverted repeats found at the termini of the M-MuLV right (R) and left (L) LTRs. L– is 5'-AATTCGTGGGGTCTTTCATT-3'; L+ is 5'-AATGAAAGACCCACG-3'; R+ is 5'-AATTCGGGGTCTTTCATT-3'; R– is 5'-AATGAAAGACCCCGG-3'. dsDNA probes were produced by mixing 5 ng of a 5'-end-labeled oligonucleotide with a sixfold excess of the complementary oligonucleotide in a 15- μ l annealing mixture containing 50 mM Tris hydrochloride (pH 7.5), 50 mM NaCl, 5 mM MgCl₂, and 0.1 mg of bovine serum albumin per ml, followed by successive 15-min incubations at 65, 37, and 20°C. 5'-

* Corresponding author.

FIG. 1. Cloning of the M-MuLV IN protein coding domain. A 2.8-kilobase *SalI*-*Bam*HI fragment of p8.8 (a full-length clone of the M-MuLV genome) was ligated into the polylinker of plasmid pTZ19R (Pharmacia). A total of 903 bases (903b) of vector and reverse transcriptase sequences were deleted by oligonucleotide-directed mutagenesis, resulting in a construct in which the IN coding domain is preceded by the first three codons of *lacZ* as shown. The sequence of the 33-base oligonucleotide used to loop out the 903 bases is shown base paired with the ss form of the initial construct. The predicted amino acid sequence of the resulting protein is indicated. PR and RT indicate the protease and reverse transcriptase coding domains of the *pol* gene, respectively.

End-labeled L⁺ was annealed with a sixfold excess of unlabeled L⁺ to produce a duplex oligonucleotide, IR_L, which corresponds to the left LTR IR. 5'-End-labeled R⁺ and a sixfold excess of unlabeled R⁺ were annealed to yield IR_R, which corresponds to the right LTR IR. 5'-End-labeled oligonucleotide SL⁺ (5'-AATTCGTGGGGTCTTTCA-3') was similarly annealed with an excess of L⁺ to produce IR_L⁺, which corresponds to the IR of the left LTR lacking the terminal two 3' thymidine residues from the proviral minus strand. Each of these dsDNA probes has a four-base overhang which represents an *Eco*RI endonuclease cleavage product. A dsDNA control probe (C) with a similar four-base overhang was produced by annealing the 5'-end-labeled oligonucleotide T1 (5'-CGATAACTGGGCCAGCTGT-3') with a sixfold excess of T2 (5'-ACAGCTGGGCCAGTT-3'). To produce ssDNA probes of specific activity equal to IR_R and C (referred to as ssIR_R and ssC), a sixfold excess of unlabeled R⁺ and T1 oligonucleotides was substituted for the complementary strands and the annealing procedure was carried out as before. The products of the annealing procedure were analyzed by electrophoresis in 20% nondenaturing polyacrylamide gels (19:1 acrylamide-bis) containing ×0.5 TBE.

Synthetic oligonucleotides were also produced for use in site-directed mutagenesis (5'-ACAGCTATGACCATGATAGAAAATTCATCACCC-3') and for sequencing of the resulting product (5'-TCATTAGGCACCCCAGGC-3').

Plasmid construction. By comparison of the published nucleotide sequence of the M-MuLV genome (29) with the results of carboxypeptidase Y degradation of reverse transcriptase (5), the IN coding domain was inferred to begin with the Ile codon beginning at nucleotide 4611. A full-length clone of the M-MuLV genome (p8.8) (30) was digested with *SalI* and *Bam*HI. A 2.8-kilobase fragment containing the IN coding domain flanked by part of the reverse transcriptase and envelope coding sequences was ligated into the poly-

linker of pTZ19R (Pharmacia) (Fig. 1). Oligonucleotide-directed mutagenesis employing a ss template produced in a *dut ung* mutant *E. coli* strain (21) was used to delete 903 base pairs of vector and reverse transcriptase sequence resulting in a construct, pIF, in which the first three codons of *lacZ* are followed by the M-MuLV IN coding domain. To confirm the identity of the deletion mutant, the resulting DNA was sequenced by using the commercial Sequenase kit (U.S. Biochemicals Corp.).

A second plasmid, pJC99, was constructed by ligating the *trpE* leader sequence of pATH3 (obtained from T. J. Kerner) to the 3' end of the IN coding domain beginning at the *Hind*III site at nucleotide 4894. The fusion protein encoded by pJC99 (TRPE-IN) was eluted from a sodium dodecyl sulfate (SDS)-polyacrylamide gel and used to produce IN-specific antibody in a New Zealand White rabbit (19).

Protein analysis: Western blots (immunoblots) and Southwestern (DNA-protein) blots. To prepare crude bacterial extracts, pelleted cells from late-logarithmic-phase cultures were boiled for 3 min in cracking buffer (10 mM sodium phosphate [pH 7.2], 6 M urea, 1% SDS, 1% 2-mercaptoethanol) and were frozen on dry ice. These extracts were stored at -70°C until used. They were subjected to electrophoresis in SDS-12.5% polyacrylamide gels (8) and either stained with Coomassie blue or transferred electrophoretically to nitrocellulose (1). Western blot analysis was performed by blocking the filters with 2% nonfat dry milk, followed by incubation with the rabbit antibody described above, and developed with an alkaline phosphatase-conjugated goat anti-rabbit antibody kit purchased from Bio-Rad Laboratories.

Southwestern analysis of DNA binding was performed essentially as described by Roth et al. (27), but labeled synthetic oligonucleotides were used as probes instead of radiolabeled plasmid DNA. Extracts of DH5α(pIF) and DH5α(pTZ19R) were subjected to electrophoresis in alternate lanes of an SDS-polyacrylamide gel and were transferred electrophoretically to a nitrocellulose filter. The filter was soaked for 1 h at 20°C in 50 ml of denaturing solution (7 M guanidine-HCl, 35 mM dithiothreitol, 50 mM Tris hydrochloride [pH 8.3], 2 mM EDTA) and washed for 1 h at 20°C in 60 to 70 ml of dilution buffer (glycerol [10% vol/vol], 0.1% Nonidet P-40, 0.5 M NaCl, 50 mM Tris hydrochloride [pH 7.5], 2 mM EDTA, 2 mM dithiothreitol). The filter was transferred to a fresh tray containing another 50 to 70 ml of dilution buffer and was stored at 4°C for a minimum of 24 h (equivalent results were seen with filters stored in this fashion for up to 8 days). The filters were placed in blocking solution (30 mM HEPES-NaOH [pH 7.6], 0.2% nonfat dry milk) for 1 h at 20°C and then were washed for a minimum of 30 min in 50 ml of incubation buffer (50 mM NaCl, 30 mM HEPES-NaOH [pH 7.6], 5 mM MgCl₂, 1 mM dithiothreitol). Duplicate segments of the filter (each with a control lane next to a lane containing the IN protein) were cut out with a razor blade and placed in sealed plastic bags containing 4 ml of incubation buffer and radiolabeled probes of equal specific activity, as determined by Cerenkov counting (1 to 3 ng of DNA with a specific activity of approximately 10⁸ cpm/μg). After incubation with mild agitation at 20°C for 1 to 3 h, the filter segments were washed separately for 30 min in 50 ml of incubation buffer, dried, and subjected to autoradiography.

In some cases, the filters were subjected to Western blot analysis as described above to verify that comparable amounts of the IN protein were present on each of the filter segments. In other experiments, the amount of DNA bound

FIG. 2. Expression of the M-MuLV IN coding domain in *E. coli* DH5 α . Cultures of DH5 α transformed with the plasmid encoding the M-MuLV IN protein, pIF, or the parent vector, pTZ19R, were grown to mid-log phase. Crude extracts of bacterial cells were prepared and subjected to electrophoresis in SDS-12.5% polyacrylamide gels. Lane 1, DH5 α (pTZ19R); lane 2, DH5 α (pIF) (culture grown without isopropyl- β -D-thiogalactopyranoside induction); lane 3, DH5 α (pIF) (culture grown with 2 mM isopropyl- β -D-thiogalactopyranoside induction); lane 4, sucrose gradient-purified M-MuLV particles disrupted in loading buffer. (A) Coomassie blue-stained gel; (B) Western blot analysis. Proteins in a gel identical to that shown in panel A were transferred electrophoretically to a nitrocellulose filter and subjected to Western blot analysis with rabbit antiserum directed against the carboxy portion of the M-MuLV IN protein. The positions of molecular size standards migrating in adjacent lanes are indicated.

to the filter fragments was measured by scintillation counting.

RESULTS

Expression of the M-MuLV IN protein in *E. coli*. Oligonucleotide-directed mutagenesis was used to construct the recombinant plasmid pIF in which the first three codons of *lacZ* are followed by the IN protein coding sequence (see Materials and Methods; Fig. 1). Direct sequence analysis of the recombinant plasmid DNA by the dideoxy method (1) confirmed that the construct had the correct nucleotide sequence (data not shown).

Crude extracts of cultures of *E. coli* DH5 α containing either pIF or the parent vector, pTZ19R, were subjected to SDS-polyacrylamide gel electrophoresis. Extracts of DH5 α (pIF) grown with or without isopropyl- β -D-thiogalactopyranoside induction (Fig. 2A, lanes 2 and 3) contained increased amounts of a protein with a molecular weight of approximately 43,000 compared with control extracts of DH5 α (pTZ19R) (Fig. 2A, lane 1). Immunoblots with rabbit antiserum elicited against a TRPE-IN fusion protein showed that the 43-kilodalton protein is the IN protein and that it migrates with the same mobility as the IN protein present in sucrose gradient-purified M-MuLV particles (Fig. 2B). The Western blots also revealed that a smaller amount of a slightly lower-molecular-weight immunoreactive species was present in the bacterial extracts (Fig. 2B; see Fig. 5B). Inspection of the published nucleotide sequence of M-MuLV (29) reveals that the first methionine codon of the IN coding region (nucleotides 4770 to 4772) is preceded by a fortuitous Shine-Dalgarno sequence that could direct translation at a site downstream from the *lac* initiation site. This second band may therefore represent the product of translational initiation at this downstream methionine codon or alterna-

FIG. 3. Oligonucleotides used to probe protein blots. Synthetic DNAs were annealed to produce duplex oligonucleotides corresponding to the IRs found at the termini of the LTRs of viral DNA (IR_L and IR_R). IR_L corresponds to the DNA sequence found at the tip of the left LTR of the putative immediate precursor to integration. Oligonucleotides lacking similarity to the tips of the M-MuLV IRs were annealed to produce C.

tively may result from proteolytic cleavage of the full-length product encoded by pIF. A precedent exists for the former possibility; Hizi and Hughes (15) found that translation directed by a similar fortuitous Shine-Dalgarno sequence present in the human immunodeficiency virus IN coding domain complicated their initial efforts to clone and express the human immunodeficiency virus IN protein in *E. coli*.

The M-MuLV IN protein expressed in *E. coli* exhibits sequence-specific binding of DNA. We looked for evidence of sequence-specific binding of DNA by the M-MuLV IN protein by using a Southwestern assay similar to that described by Roth et al. (26) but differing principally in the use of radioactively labeled synthetic oligonucleotides (Fig. 3) as probes instead of plasmid DNA. dsDNA probes were produced by annealing the 5'-end-labeled oligonucleotide with a sixfold excess of the complementary oligonucleotide; this ratio was chosen to ensure that all of the labeled oligonucleotide would be base paired to its complement after the annealing procedure. Electrophoresis of the annealed oligonucleotides in a nondenaturing polyacrylamide gel was used to examine the completeness of the annealing reaction (Fig.

FIG. 4. Electrophoretic analysis of oligonucleotide probes. Oligonucleotides L⁻, SL⁻, R⁺, and T1 were radiolabeled with T4 kinase and [γ -³²P]ATP and annealed with a sixfold excess of either the same or a complementary unlabeled oligonucleotide to yield ssDNA and dsDNA probes of equal specific activity. These probes were analyzed by electrophoresis through a 20% nondenaturing polyacrylamide gel. Lanes 1, 4, 7, and 10, radiolabeled oligonucleotides R⁺, L⁻, SL⁻, and T1 (unannealed controls), respectively; lanes 2, 5, 8, and 11, ss probes ssIR_R, ssIR_L, ssIR_L*, and ssC, respectively; lanes 3, 6, 9, and 12, ds probes IR_R, IR_L, IR_L*, and C, respectively. Overexposure of this and other autoradiograms demonstrated that after incubation with its complementary strand, essentially all of the radiolabeled oligonucleotide was present as a slower migrating, presumably duplex, species.

FIG. 5. Sequence-specific DNA binding activity of the M-MuLV IN protein expressed in *E. coli*. (A) Proteins in crude extracts of DH5 α (pTZ19R) and DH5 α (pIF) were subjected to electrophoresis in SDS-polyacrylamide gels, transferred to nitrocellulose, renatured, and probed with 32 P-labeled single-stranded (SS) or double-stranded (DS) oligonucleotides corresponding to the M-MuLV LTR right IR or to C. After being washed to remove unbound probe, the filters were subjected to autoradiography. Lanes 1, 3, 5, and 7, extracts of DH5 α (pIF); lanes 2, 4, 6, and 8, extracts of DH5 α (pTZ19R). (B) After autoradiography, the filter segments were subjected to Western blot analysis.

4). The shift in the mobility of each of the labeled oligonucleotides after annealing with the complementary oligonucleotide is consistent with the formation of a duplex structure. The lack of residual material at the position of the ss oligonucleotide indicates that the annealing had gone to completion.

Extracts of DH5 α (pIF) and DH5 α (pTZ19R) were run side by side in an SDS-polyacrylamide gel, transferred to nitrocellulose, denatured in guanidine, and renatured. Duplicate portions of a single filter, each containing the side-by-side extracts described above, were used in each experiment. They were incubated with ss or ds synthetic oligonucleotides corresponding to the IR_R of the M-MuLV right LTR or to the unrelated control oligonucleotide (C) (Fig. 3). After incubation, the filters were washed briefly to remove unbound DNA and were subjected to autoradiography. In some experiments, autoradiography of the filters was followed by immunoblot analysis to verify that lanes on the blot contained comparable amounts of IN protein. No binding of the oligonucleotide probes to proteins in the control extracts from DH5 α (pTZ19R) was observed under the conditions used (even-numbered lanes, Fig. 5 and 6).

Autoradiographs of the blots consistently revealed much greater binding of the dsIR_R probe to the IN protein as compared with the binding of the control probe (Fig. 5A, lanes 1 and 3). This result was confirmed by cutting out the bands and quantifying the bound label by scintillation counting. The Western blot analysis (Fig. 5B, lanes 1 and 3) confirmed that comparable amounts of the IN protein were present in each of the lanes containing the DH5 α (pIF) extract. The binding of IR_R does not depend on the presence of MgCl₂; comparable binding was observed in experiments

FIG. 6. DNA binding specificity of the M-MuLV IN protein expressed in *E. coli*. Protein blots were prepared as described in the legend to Fig. 5. Lanes 1, 3, 5, and 7, DH5 α (pIF) extracts; lanes 2, 4, 6, and 8, DH5 α (pTZ19R) extracts. Separate segments of the blot were incubated with 32 P-labeled ds oligonucleotide probes as follows. Lanes 1 and 2, IR_L; lanes 3 and 4, IR_L+; lanes 5 and 6, IR_R; lanes 7 and 8, C. After being washed to remove unbound probe, the filters were subjected to autoradiography.

in which the incubation buffer contained 2 mM EDTA (data not shown). Sequence-specific DNA binding was also seen when filters were incubated with ss oligonucleotides; the ssIR_R probe was bound to a greater extent than was the ssC probe (Fig. 5A, lanes 5 and 7). With the ss oligonucleotides, the difference between the IR and control probes was not as striking as that seen with ds probes but was consistently observed by both autoradiography and scintillation counting of the filters. A similar preference for the ss form of the IR oligonucleotide was observed when the binding of ssIR_R was compared with that of other ss oligonucleotides lacking any sequence similarity to the M-MuLV LTR IRs (data not shown).

Comparison of binding of full-length and processed IR oligonucleotides by the M-MuLV protein. Within the first 24 h of infection, the retroviral reverse transcriptase synthesizes a linear dsDNA intermediate with the RNA genome as a template. Recent data suggest that this intermediate is subsequently processed for integration by the removal of two bases (thymidine residues) from the 3' ends of the IRs at the termini of the LTRs (3, 12, 27). Southwestern blots, prepared as described above, were incubated with the dsDNA oligonucleotide probes corresponding to IR_L and IR_R and, in addition, to a probe lacking the two 3' thymidine kinase residues (IR_L+). IR_R and IR_L were bound to a much greater extent than was C (Fig. 6). In the experiment shown, IR_L+ was bound poorly compared with the full-length inverted repeat probes, with the binding being comparable to that of C. However, in other experiments, IR_L+ was bound to nearly the same extent as were IR_R and IR_L. This variation in the binding of IR_L+ did not appear to be due to differences in the amount of the cloned IN protein transferred to the nitrocellulose filters or in the specific activity of the probes used. It is possible that slight variations in the extent of protein renaturation on the filter could account for the variation in binding from experiment to experiment (see Discussion).

DISCUSSION

During retroviral replication, reverse transcriptase uses the RNA genome as a template to synthesize a linear dsDNA intermediate that is subsequently integrated into the host genome in a process involving the viral IN protein. The viral coding regions of this DNA intermediate are flanked by LTRs and the tips of the LTRs contain IR sequences vital to the integration process (4, 27).

In this study, we have presented evidence for sequence-specific DNA binding by the M-MuLV IN protein by using oligonucleotide probes in a Southwestern assay. While synthetic oligonucleotides have been extensively used to study DNA-protein binding by gel mobility retardation (32) and to purify DNA-binding proteins by affinity chromatography (17), we know of no precedent for their use in this fashion; in previous descriptions of the so-called Southwestern technique, radiolabeled plasmid DNA was used to probe protein blots (23, 26). Southwestern analysis with oligonucleotide probes could be very useful in studying the DNA binding activity of proteins present in crude extracts or when the physical properties of a protein preclude other means of analysis.

Roth et al. (26) studied the binding of plasmid DNA probes by a series of TRPE-IN fusion proteins by using a similar Southwestern assay. They found no evidence for preferential binding of large DNA fragments containing the palindromic LTR-LTR circle junction by a protein composed of the complete IN amino acid sequence preceded by 26 foreign amino acids. There are several possible explanations for the differences between their findings and the results presented here. First, the foreign amino acid residues added to the N terminus of the IN protein may have altered its tertiary structure and thereby its binding specificity. In contrast, we cloned the M-MuLV IN coding domain as a fusion gene preceded by only three codons of *lacZ*. In *E. coli*, the initial *N*-formylmethionine of nascent proteins is frequently removed after translation if followed by a threonine residue (31). Thus, the bacterial extracts used in the Western and Southwestern assays described here probably contain a protein in which the IN amino acid sequence is preceded by at most two additional amino acids (Thr-Met, Fig. 1), a more authentic representation of the viral protein than that used by Roth et al. (26). Second, in their experiments, Roth et al. (26) examined the binding of the viral LTR-LTR circle junction sequence cloned as a part of a much larger plasmid. It is likely that, in addition to binding with a high specificity to a relatively short target sequence, the IN protein also binds nonspecifically to DNA with a lower affinity. If this is the case, then the vast excess of vector sequences over specific target sequences in the probe may have precluded detection of the specific component of the binding reaction. Thus, a control plasmid containing no viral nucleotide sequences might be bound to a similar extent in the Southwestern assay as a plasmid containing the LTR-LTR circle junction. Similar phenomenon are well described; bacteriophage lambda integrase (18), *E. coli* RNA polymerase (14), and a number of repressor proteins all bind specifically to their respective target sequences yet exhibit significant non-specific DNA binding as well. Third, recent data suggest that the circularized viral DNA containing the LTR-LTR junction sequence is inefficiently utilized as a precursor to the integrated provirus (2, 10, 22). It is therefore possible that Roth et al. (26) found no evidence for specific binding of the circle junction simply because the binding activity of the IN protein requires that the IR target sequence be present at the terminus of a linear DNA molecule.

Retroviral integration depends on the presence of the IN protein (6, 7, 27, 28) and the IRs present at the termini of the LTRs (4). Our results showing binding of the IR sequence by the IN protein demonstrate a specific interaction between these elements. Recent data suggest that the initial step of integration involves the processing of the linear viral DNA by removal of the terminal two bases from the 3' ends of the IR (3, 12, 28). There is certainly reason to suspect that the

M-MuLV IN protein itself might be an endonuclease and be responsible for this processing. An endonuclease activity has been found in purified viral cores of Rauscher MuLV (20). The corresponding *pol* gene product, pp32, from avian sarcoma and leukemia viruses has been shown to bind the viral LTRs and cleave viral DNA near the site of integration (9, 13, 24). However, Panet and Baltimore (25) have presented evidence that the major endonuclease activity found in M-MuLV virions is probably not the product of the IN coding region.

In related studies, the IR oligonucleotides used in this study have been incubated with detergent-disrupted M-MuLV virions and the 3' ends corresponding to the tips of the LTRs have been found to be shortened by two bases (L. Ishimoto, M. Halperin, and J. Champoux; unpublished data). This reaction appears to mimic the processing of viral DNA, which occurs prior to integration. However, we have been unsuccessful in attempts to demonstrate any endonuclease activity associated with cloned M-MuLV IN protein that had been eluted from an SDS-polyacrylamide gel and renatured in vitro. Roth et al. (26) were also unable to demonstrate endonuclease activity associated with their TRPE-IN fusion proteins. It is possible that the M-MuLV IN protein actually has an endonuclease activity which has not yet been demonstrated for technical reasons. However, the processing of LTR ends prior to integration might instead involve a cellular endonuclease; the enzyme detected by Panet and Baltimore (25) in M-MuLV IN-deficient viral particles is a possible candidate. If a cellular endonuclease is in fact involved, then perhaps the IN protein acts as a kind of template for a cellular nuclease that removes the two bases from the 3' ends of the viral DNA.

We sought to determine whether IR_L*, an oligonucleotide probe corresponding to the termini of the putative immediate precursor to the integrated provirus (i.e., lacking the two 3' terminal T residues), would be bound by the IN protein. Thus far, these experiments have yielded variable results. Since we never observed any variations from experiment to experiment with the full-length oligonucleotide probes, we conclude that the interaction of the IN protein with the shortened probe in a Southwestern assay is somehow different from the interaction with the full-length duplex oligonucleotide. This result might be explained by assuming that the binding of the shortened oligonucleotide is slightly weaker than the binding of the full-length probe and that there is some variation from experiment to experiment in the extent of IN protein renaturation during the preparation of the filters for the Southwestern assays. Accordingly, partially renatured as well as fully renatured IN protein may be equally capable of specifically binding the full-length probe, whereas any partially renatured protein, when present, would be unable to bind the shortened oligonucleotide. Further experiments will be required to elucidate the basis for these observations.

ACKNOWLEDGMENTS

This work was supported by grants from the National Institute of Health (CA51605) and the American Cancer Society (MV-232C). P.A.K. was supported by training grant T32-HD07233 from the National Institutes of Health.

The authors wish to thank Alison Rattray for advice and technical assistance and Lance Ishimoto for helpful suggestions throughout the work.

LITERATURE CITED

1. Ausubel, F., R. Brent, R. Kingston, D. Moore, J. Seidman, J. Smith, and K. Struhl (ed.). 1987. Current protocols in molecular biology, chapters 1 and 10. John Wiley & Sons, Inc., New York.
2. Brown, P., B. Bowerman, H. Varmus, and J. M. Bishop. 1987. Correct integration of retroviral DNA in vitro. *Cell* **49**:347-356.
3. Brown, P., B. Bowerman, H. Varmus, and J. M. Bishop. 1989. Retroviral integration: structure of the initial covalent product and its precursor, and a role for the IN protein. *Proc. Natl. Acad. Sci. USA* **86**:2525-2529.
4. Colicelli, J., and S. Goff. 1988. Sequence and spacing requirements of a retrovirus integration site. *J. Mol. Biol.* **199**:47-59.
5. Copeland, T. D., G. F. Gerard, C. W. Hixson, and S. Oroszlan. 1985. Amino- and carboxy-terminus sequence of Moloney murine leukemia virus reverse transcriptase. *Virology* **143**:676-679.
6. Donehower, L. 1989. Analysis of mutant Moloney murine leukemia viruses containing linker insertion mutations in the 3' region of *pol*. *J. Virol.* **62**:3958-3964.
7. Donehower, L., and H. 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.
8. Dreyfuss, G., S. Adam, and Y. D. Choi. 1984. Physical changes in cytoplasmic messenger ribonucleoproteins in cells treated with inhibitors of mRNA transcription. *Mol. Cell. Biol.* **4**:415-423.
9. Duyk, G., J. Leis, M. Longiaru, and A. 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.
10. Ellis, J., and A. Bernstein. 1989. Retrovirus vectors containing an internal attachment site: evidence that circles are not intermediates to murine retrovirus integration. *J. Virol.* **63**:2844-2846.
11. Fujiwara, T., and R. Craigie. 1989. Integration of mini-retroviral DNA: a cell-free reaction for biochemical analysis of integration. *Proc. Natl. Acad. Sci. USA* **86**:3065-3069.
12. Fujiwara, T., and K. Mizuuchi. 1988. Retroviral DNA integration: structure of an integration intermediate. *Cell* **54**:497-504.
13. Grandgenett, D., A. C. Vora, and R. Schiff. 1978. A 32,000-dalton nucleic acid-binding protein from avian retrovirus cores possesses DNA endonuclease activity. *Virology* **89**:119-132.
14. Hinkle, D. C., and M. J. Chamberlin. 1972. Studies on the binding of *Escherichia coli* RNA polymerase to DNA. I. The role of sigma subunits in site selection. *J. Mol. Biol.* **70**:157-185.
15. Hizi, A., and S. Hughes. 1988. Expression of the Moloney murine leukemia virus and human immunodeficiency virus integration proteins in *Escherichia coli*. *Virology* **167**:634-638.
16. Hu, S., D. Court, M. Zweig, and J. 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.
17. Kadonga, J. T., and R. Tjian. 1986. Affinity purification of sequence-specific DNA binding proteins. *Proc. Natl. Acad. Sci. USA* **83**:5889-5893.
18. Kikuchi, Y., and H. A. Nash. 1979. Nicking-closing activity associated with bacteriophage lambda *int* gene product. *Proc. Natl. Acad. Sci. USA* **76**:3760-3764.
19. Konopka, J. B., R. L. Davis, S. M. Watanabe, A. S. Ponticelli, L. Schiff-Maker, N. Rosenberg, and O. N. Witte. 1984. Only site-directed antibodies reactive with the highly conserved *src*-homologous region of the *v-abl* protein neutralize kinase activity. *J. Virol.* **51**:223-232.
20. Kopchick, J., J. Harless, B. Geisser, R. Killam, R. Hewitt, and R. Arlinghaus. 1981. Endodeoxynuclease activity associated with Rauscher murine leukemia virus. *J. Virol.* **37**:274-283.
21. Kunkel, T., J. D. Roberts, and R. A. Zankour. 1987. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol.* **154**:367-382.
22. Lobel, L., J. Murphy, and S. Goff. 1989. The palindromic LTR-LTR junction of Moloney murine leukemia virus is not an efficient substrate for proviral integration. *J. Virol.* **63**:2629-2637.
23. Miskimins, W. K., M. Roberts, A. McClelland, and F. Ruddle. 1985. Use of protein-blotting procedure and a specific DNA probe to identify nuclear proteins that recognize the promoter of the transferrin receptor gene. *Proc. Natl. Acad. Sci. USA* **82**:6741-6744.
24. Misra, T. K., D. P. Grandgenett, and J. T. Parsons. 1982. Avian retrovirus pp32 DNA-binding protein. I. Recognition of specific sequences in retrovirus DNA terminal repeats. *J. Virol.* **44**:330-343.
25. Panet, A., and D. Baltimore. 1987. Characterization of endodeoxynuclease activities in Moloney murine leukemia virus and its replication deficient mutants. *J. Virol.* **61**:1756-1760.
26. Roth, M., N. Tanese, and S. Goff. 1988. Gene product of Moloney murine leukemia virus required for proviral integration is a DNA-binding protein. *J. Mol. Biol.* **203**:131-139.
27. Roth, M. J., P. Schwartzberg, and S. 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.
28. Schwartzberg, P., J. Colicelli, and S. 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.
29. Shinnick, T., R. Lerner, and J. Sutcliffe. 1981. Nucleotide sequence of Moloney murine leukemia virus. *Nature (London)* **293**:543-548.
30. Shoemaker, C. S., S. Goff, E. Gilboa, M. Pashind, 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.
31. Tsunasawa, J., J. W. Stewart, and F. S. Sherman. 1985. Amino terminal processing of mutant forms of yeast iso-1-cytochrome c: the specificities of methionine aminopeptidase and acetyltransferase. *J. Biol. Chem.* **260**:5382-5391.
32. Turner, R., and R. Tjian. 1989. Leucine repeats and an adjacent DNA binding domain mediate the formation function cFos and cJun heterodimers. *Science* **243**:1689-1694.
33. Varmus, H., and R. Swanstrom. 1982. Replication of retroviruses. p. 369-512. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), *Molecular biology of tumor viruses: RNA tumor viruses*, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.