JOURNAL OF VIROLOGY, Mar. 1991, p. 1160–1167
0022-538X/91/031160-08$02.00/0

Vol. 65, No. 3

# Defining Nucleic Acid-Binding Properties of Avian Retrovirus Integrase by Deletion Analysis

STEVEN R. MUMM* AND DUANE P. GRANDGENETT

*Institute for Molecular Virology, St. Louis University Medical Center, 3681 Park Avenue, St. Louis, Missouri 63110*

Received 13 December 1989/Accepted 3 December 1990

**Integration of retroviral DNA into the host genome requires the activity of retrovirus-encoded integration protein IN. We expressed Rous sarcoma virus (RSV) IN, 286 amino acid residues in length, by using in vitro transcription, followed by in vitro translation in rabbit reticulocyte lysate. The nucleic acid-binding activity of in vitro-translated IN was assessed by using DNA-cellulose affinity chromatography and poly(U)-Sepharose affinity chromatography and by sedimentation analysis in the presence or absence of DNA. In vitro-translated RSV IN exhibited nucleic acid-binding activity similar to that of IN purified from avian myeloblastosis virus. To identify regions of IN which bind to nucleic acids, several deletions of RSV IN were generated. The $NH_2$-terminal 26 amino acids, including the two His residues of a His-Cys box, were not necessary for IN nucleic acid binding with any of the substrates tested. The substrates included native calf thymus DNA, poly(U), and a double-stranded linear DNA molecule with RSV long terminal repeat sequences at its termini. The COOH-terminal region (residues 178 to 286) of IN bound quantitatively (greater than 90%) to poly(U) and to single-stranded circular φX174 DNA but did not exhibit the double-stranded linear DNA-binding ability of the entire IN molecule.**

An important step in retrovirus replication is integration of viral DNA into the host genome. The 3′ end of the retrovirus *pol* gene encodes integration protein IN. The function of this protein was established by generating and analyzing missense mutations in the 3′ end of the *pol* genes of various retrovirus species (6, 34, 35, 39). These mutant viruses were able to generate viral DNA through reverse transcription but were unable to integrate their viral DNAs into the host genome efficiently. Further analysis revealed that reverse transcription generates linear blunt-ended, double-stranded (ds) viral DNA (4, 10, 36). IN is responsible for removing two nucleotides from each 3′ end of the linear viral DNA (4, 36). The 3′ OH recessed ends of the viral DNA are then coupled to 5′ ends of the host DNA. Genetic evidence has not proven that IN is responsible for nicking of the host DNA or joining of the viral DNA to the host DNA, although baculovirus-expressed Moloney murine leukemia virus (M-MuLV) IN is capable of joining viruslike DNA molecules in vitro (5).

Of all the retrovirus IN proteins, that of avian retroviruses has been the most thoroughly characterized, largely by use of purified avian myeloblastosis virus (AMV) IN (pp32) (15, 27, 42). The properties associated with avian retroviral IN include DNA-binding, RNA-binding, and DNA endonuclease activities. Purified AMV IN can remove the correct nucleotides from the 3′ ends of long terminal repeat (LTR) viral DNA substrates, mimicking the in vivo action of IN (26, 43). AMV IN can also generate 5′ overhang nicks in a supercoiled DNA substrate at the LTR-LTR junction, suggesting that IN is responsible for generating 5′ overhang nicks in the host DNA in vivo (16, 42).

AMV IN shows some sequence specificity in DNA binding. The nitrocellulose filter DNA-binding assay shows that AMV IN preferentially binds viral LTR sequences over *gag* or *src* DNA sequences (27). However, AMV IN also preferentially binds to or nicks certain pBR322 DNA sequences, including fragments containing the *amp* gene promoter and other A+T-rich fragments (27, 42). AMV IN preferentially binds to supercoiled and single-stranded (ss) DNA over ds linear DNA (27). M-MuLV IN expressed in *Escherichia coli* also preferentially binds to ss DNA over ds DNA and has no apparent binding preference for the LTR-LTR junction region (37). Recently, ss and ds oligonucleotides corresponding to the termini of unintegrated linear viral DNA were found to be specifically bound by bacterially expressed M-MuLV IN (28).

AMV IN binds to various nucleic acid affinity chromatography media and has been purified by using these techniques. For these reasons, we decided initially to identify regions of IN involved in DNA binding by using nucleic acid affinity chromatography. We used in vitro transcription and translation to express the Rous sarcoma virus (RSV) IN protein and various deletion mutants of IN. These proteins were tested for DNA and RNA binding by nucleic acid affinity chromatography. The results were confirmed by sedimentation analysis in the presence or absence of DNA (7, 40). In vitro-translated IN has DNA- and RNA-binding activities similar to those of AMV IN. Deletion of the $NH_2$-terminal 26 amino acids of IN, including two His residues of a His-Cys box, had little effect on nucleic acid binding. The COOH-terminal half of IN was required for nucleic acid binding and was sufficient for binding to ss nucleic acids.

## MATERIALS AND METHODS

**Construction of vectors.** Plasmid pG36 was constructed by inserting the 1,328-bp *Bgl*II-to-*Pst*I fragment from the 3′ end of *pol* (RSV Prague A strain [PrA]) (19) into pGEM-4 (Promega Biotec) (see Fig. 1). Synthetic nucleotides for both strands were used to reconstruct the 5′ end of the gene for IN (19). An *Asp*718 site joining the 5′ end of the synthetic nucleotide to the vector was later destroyed by partial

* Corresponding author.

digestion with *Asp*718, filling in of the ends with Klenow, and ligation. Plasmid pG32 was generated by introducing a stop codon into pG36 (Fig. 1) by oligonucleotide mutagenesis with the Bio-Rad Muta-Gene kit. While RSV PrA DNA was used in all of the cloning steps, the nucleotide sequence of the closely related Prague C (PrC) strain of RSV was used to predict the amino acid sequence (38). The 5′ end of the RSV PrA gene for p36/p32 (including the synthetic nucleotide) was sequenced down to the Met codon corresponding to amino acid 27. The PrA sequence did not vary from the published sequence of PrC. After generation of the stop codon in pG32, the 3′ end of the gene for IN was sequenced from nucleotides 4995 to 5152 (nucleotide numbers correspond to the PrC sequence), confirming that the stop codon was generated properly. One nucleotide difference was detected between the PrA and PrC strains. Nucleotide 5010 is G in PrA (reported as A in PrC; 38). This nucleotide difference does not affect the predicted amino acid sequence.

Plasmid pG32Bgl was constructed by insertion of the *Bgl*II-to-*Pst*I fragment from pG32 containing the gene for IN into pGEM-4 at the *Bam*HI and *Pst*I sites. Plasmid pG32C-term was constructed by cutting pG36 with *Nru*I and treating it with BAL 31 for 120 s to remove an out-of-frame start codon. The plasmid was further cut with *Eco*RI, filled in with Klenow, and ligated. The p32 stop codon was then inserted by using the *Kpn*I-to-*Xho*I fragment. Proper construction of pG32C-term was substantiated by DNA sequencing. One nucleotide difference between the PrA and PrC strains was detected in nucleotides 4634 to 4753. Nucleotide 4715 is A in PrA and has been reported as G in PrC (38). This nucleotide difference is responsible for loss of the *Bam*HI site in PrA and a conservative amino acid change (Arg [PrC] to Lys [PrA]).

**Expression of p32-related proteins.** Ten micrograms of pG36 was linearized with *Xho*I and transcribed with SP6 polymerase by using the Riboprobe Gemini System (Promega Biotec) as recommended by the supplier. The reaction mixture was treated with RNase-free DNase (Cooper), phenol extracted, and ethanol precipitated. Plasmids pG32, pG32Bgl, and pG32C-term were treated in the same manner. To generate mRNA for $NH_2$-terminal products of p32, pG32 was digested with *Dde*I, *Nru*I, or *Asp*718 and transcribed as described above. Usually, 2 μg of mRNA, reconstituted in diethyl pyrocarbonate-treated water, was translated in vitro by using a rabbit reticulocyte system (Promega Biotec) as recommended by the supplier. The lysates were then made 2 mM in phenylmethylsulfonyl fluoride (PMSF) and incubated for 15 min with 20 μg of heat-treated RNase A (Calbiochem).

**Immunoprecipitation.** All sera were incubated on ice with 0.1 volume of 10% Trasylol–50 mM NaCl–20 mM Tris hydrochloride (pH 7.5)–1 mM EDTA–1 mM PMSF for 30 min prior to use. Samples were made 1% sodium dodecyl sulfate (SDS)–10 mM dithiothreitol and heated for 3 min to 100°C prior to use. Samples were incubated with 5 μl of sera overnight at 4°C in 350 μl of 20 mM Tris hydrochloride (pH 7.5)–140 mM NaCl–0.5% deoxycholate–0.5% Nonidet P-40–0.1% SDS–1 mg of bovine serum albumin per ml–1% Trasylol–1 mM PMSF (buffer A). The samples were then treated with 2 mg of protein A-Sepharose (Pharmacia) in 15 μl of buffer A for 1 h at 4°C. The immune complexes were pelleted and washed twice in 20 mM Tris hydrochloride (pH 7.5)–50 mM NaCl–0.5% deoxycholate–0.5% Nonidet P-40–0.1% SDS–1 mM PMSF, twice in 20 mM Tris hydrochloride (pH 7.5)–50 mM NaCl–0.5% Nonidet P-40–0.1% SDS, and twice in 20 mM Tris hydrochloride (pH 7.5). Pellets were heated at 100°C for 5 min in 40 μl of Laemmli loading buffer and loaded onto SDS-polyacrylamide gels (29).

To generate C-2 antiserum, the *Bgl*II-to-*Xho*I DNA fragment encompassing the entire PrA IN protein, except for the first seven amino acids, was cloned into the pATH2 vector at its *Bam*HI site (41). The TrpE-IN fusion protein expressed in *E. coli* RR1 was isolated by SDS-polyacrylamide gel electrophoresis and injected into rabbits to generate antibodies (C-2). TrpE fusion vector pATH2 was the kind gift of T. J. Koerner.

**Sequencing.** [$^3$H]Leu or [$^3$H]Val (100 μCi; ICN Radiochemicals) was dried under vacuum in microcentrifuge tubes. Rabbit reticulocyte lysate (35 μl) was added, and the mixture was preincubated with 2.5 U of citrate synthase (from porcine heart) (Sigma) and 3 μl of 20 mM oxaloacetate (pH 7.0; Sigma) at 30°C for 15 min (33). Additional reagents, including [$^{35}$S]Met, were added and incubated as recommended by the supplier (Promega Biotec). The lysates were run on SDS-polyacrylamide gels. The gels were treated with 10% acetic acid for 30 min, rinsed in three changes of water for 30 min, dried, and autoradiographed. Proteins of interest were excised and eluted from the gel by diffusion in 2 ml of 25 mM Tris–192 mM glycine–0.1% SDS. Five micrograms of β-lactoglobulin (Applied Biosystems) was added as the carrier and sequencing standard. The proteins were dialyzed extensively against 0.01% SDS and lyophilized. The proteins were reconstituted in 200 μl of $H_2O$ and sequenced by automated Edman degradation with an Applied Biosystems 477A Protein Sequencer (18). Total radioactivity was determined in fractions from each sequencing cycle.

**DNA-cellulose affinity chromatography.** Native DNA-cellulose (0.6 ml) (Pharmacia) was rinsed for several hours with buffer B, which contained 20 mM Tris hydrochloride (pH 7.5) at 25°C, 50 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol, 10% glycerol, and 100 μg of bovine serum albumin (Bethesda Research Laboratories) per ml. Translation reaction mixtures were diluted 1:2 with buffer B and clarified twice for 5 min each time in an Eppendorf microcentrifuge. Samples were loaded onto the columns and allowed to bind for 15 min before being washed with several column volumes of buffer B. Samples were eluted with a 10-ml linear NaCl gradient (50 mM to 1 M). All procedures were performed at 4°C with a flow rate of 4 ml/h. The same procedure was used for poly(U)-Sepharose affinity chromatography, except that the NaCl gradient was 100 mM to 1 M.

**Sedimentation DNA-binding assay.** IN-related proteins were in vitro translated, treated with RNase, and clarified as previously described. Lysate (10 μl) was incubated with 10 μg of DNA in 180 μl of buffer (20 mM Tris hydrochloride [pH 7.5], 50 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol, 40 μg of bovine serum albumin per ml) for 1 h at 4°C (40). The reaction mixtures were sedimented on 5 to 20% sucrose gradients at 4°C. The gradients were fractionated from the bottom and analyzed by SDS-polyacrylamide gel electrophoresis and autoradiography. Fractions were also analyzed on 1% agarose gels with ethidium bromide staining. DNA substrate pFT-NdeI (8a) is a 2,931-bp pGEM-3-based vector (Promega) containing adjacent RSV LTR short inverted repeats. Digestion with *Nde*I generates a linear ds molecule with LTR short inverted repeats at the termini (9). The LTR U5 and U3 sequences were 36 and 24 bp long, respectively. This substrate was sedimented for 5 h at 44,000 rpm in an SW 50.1 rotor. ϕX174 DNA substrates were sedimented for 200 min at 44,000 rpm in an SW 50.1 rotor.

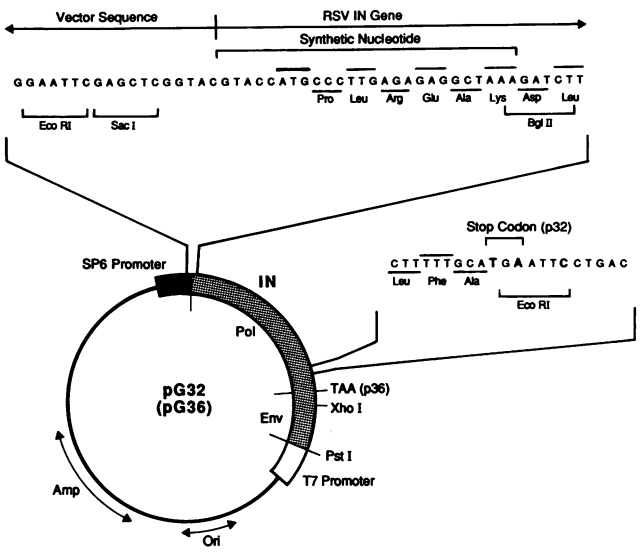


FIG. 1. Vector used for in vitro transcription from the gene for IN (PrA strain of RSV). The 3′ region of *pol* and part of *env* (stippled region) were cloned into the pGEM-4 vector downstream of the SP6 promoter. A synthetic nucleotide at the 5′ end of the gene for IN supplies an initiator Met codon and properly reconstructs the 5′ end of the gene down to the *Bgl*II site. A stop codon, required for expression of p32, was generated by oligonucleotide-directed mutagenesis. A similar vector, pG36, without the stop codon, was used for transcription of p36 mRNA. Nucleotides underlined or overlined identify codons in frame with the initiator Met codon.

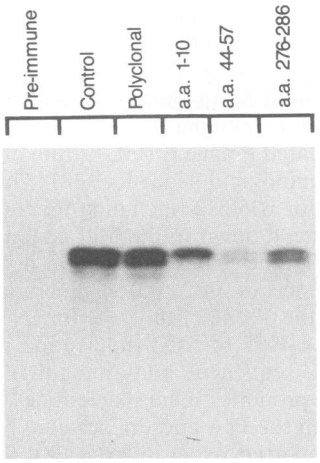


FIG. 2. Immunoprecipitation of proteins produced by in vitro translation of p36. In vitro-translated p36 was purified by poly(U)-Sepharose affinity chromatography. Approximately 40,000 cpm of [$^{35}$S]Met-labeled p36 was immunoprecipitated with preimmune rabbit serum and various antisera as described in Materials and Methods. The immunoprecipitates and control, 40,000 cpm of purified p36, were electrophoresed on an SDS–10% polyacrylamide gel and autoradiographed. Polyclonal refers to antiserum raised against an IN fusion protein described in Materials and Methods. The terms a.a. 1–10, a.a. 44–57, and a.a. 276–286 refer to antisera raised against IN peptides of the corresponding amino acid residues.

## RESULTS

**Expression of p32-related (IN) proteins.** During viral infection, the IN protein (pp32 in RSV) is produced by proteolytic processing of a *gag-pol* precursor (2, 8, 14, 19). Proteolytic processing generates both the $NH_2$ and COOH termini of IN. Hence, there is neither an initiator Met nor a stop codon for the gene for IN. The p36 protein apparently arises from incomplete processing at the COOH terminus of p32. A 4.1-kDa polypeptide is cleaved from p36 at its COOH terminus to generate p32 (see Fig. 3). For expression of p36, plasmid pG36 was constructed by inserting the 3′ end of the *pol* gene of RSV PrA into vector pGEM-4 (Fig. 1). The initiator Met was supplied by a synthetic nucleotide which properly reconstructs the 5′ end of the gene for IN. Plasmid pG36 was linearized at the *Xho*I site and transcribed with SP6 polymerase, typically resulting in 20 μg of p36-specific mRNA. Translation of p36 mRNA (2 to 3 μg) in a rabbit reticulocyte lysate system produced two major polypeptides, as shown by SDS-polyacrylamide gel analysis (Fig. 2). Immunoprecipitation with IN-specific antisera showed that the proper reading frame had been maintained (Fig. 2). Both proteins immunoprecipitated with polyclonal antiserum generated from a p32 fusion protein. Likewise, both proteins were precipitated with antisera generated from peptides corresponding to amino acids 44 to 57 and 276 to 286 (14). However, only the slower migrating band immunoprecipitated with antiserum generated from a peptide encompassing amino acids 1 to 10. This suggested that the slower migrating protein was p36 and the faster migrating protein arose from proteolytic processing of p36 or initiation of translation from an internal Met codon.

For expression of p32 (IN), a stop codon was introduced in pG36 by using site-directed oligonucleotide mutagenesis (Fig. 1). The stop codon was generated precisely at the COOH terminus of p32 (14). In vitro translation of p32 produced two major proteins with immunoreactivities similar to those of the proteins translated from pG36 (data not shown). To confirm that the slower migrating protein was p32, in vitro translation of p32 was performed with [$^3$H]Leu (33). The [$^3$H]Leu-labeled slower migrating protein was excised from an SDS-polyacrylamide gel and sequenced from its $NH_2$ terminus by automated Edman degradation (30). [$^3$H]Leu eluted in cycles 2, 8, and 12, which indicated that the slower migrating protein was p32 lacking the initiator Met (Fig. 1) (19, 38). The faster migrating protein presumably resulted from initiation of translation at amino acid 27 (see below). This internal initiation codon, as well as others (see below), permitted examination of two IN products simultaneously in our DNA-binding assays.

The $NH_2$-terminal peptides of p32 (N9.9K, N14.7K, and N28.5K; Fig. 3) were produced by digesting pG32 with *Dde*I, *Nru*I, or *Asp*718, followed by in vitro transcription with SP6 polymerase. Transcription occurs from the promoter down to the respective restriction site. Likewise, in vitro translation generates peptides with COOH termini corresponding to their respective restriction sites. SDS-polyacrylamide gel electrophoresis revealed that the $NH_2$-terminal peptides were of the proper sizes in relation to p36 and p32 (data not shown). Immunoprecipitation showed that the peptides were of the proper reading frame and initiated at the $NH_2$ terminus of p32 (data not shown).

In vitro translation of the COOH-terminal peptides relied on various deletions of the 5′ end of the gene for IN in pG32 and the use of internal Met codons for initiation of translation. Translation of C28.5K (Fig. 3) resulted from deletion of the 5′ end of IN down to the *Bgl*II site. One major protein, which comigrated with the faster migrating protein from pG32, was produced (see Fig. 5d). The identity of this protein was confirmed by using immunoprecipitation and amino acid sequencing. The C28.5K protein was translated

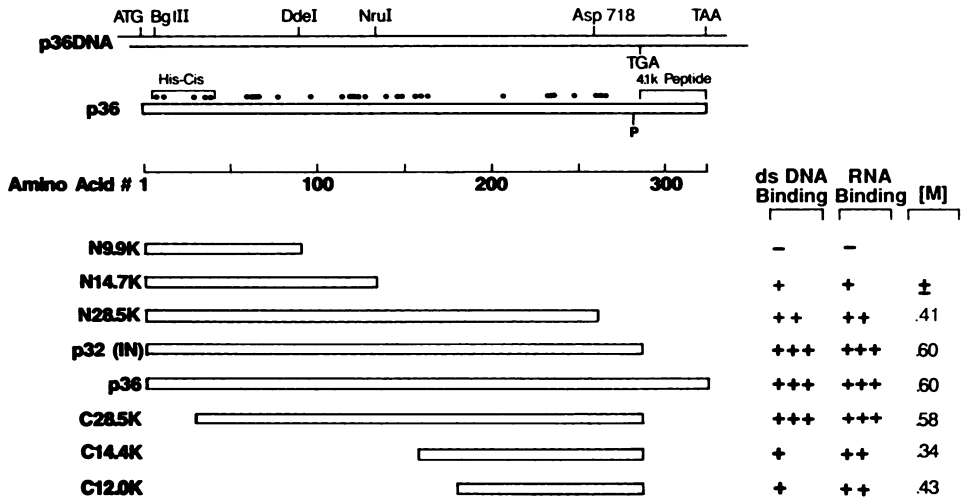


FIG. 3. Schematic representation of in vitro-translated IN-related proteins. The top line (p36 DNA) identifies restriction sites used in cloning and for in vitro transcription for $NH_2$-terminal products. The second line (p36) represents the p36 protein. The dots identify conserved amino acids among RSV, human immunodeficiency viruses 1 and 2, human T-lymphotropic virus types I and II, and simian immunodeficiency virus from an African green monkey and an Asian macaque monkey (31a). His-Cis identifies conserved His at amino acids 9 and 13 and Cys at amino acids 37 and 40 in RSV. The site of phosphorylation at Ser residue 282 in pp32 (isolated from RSV PrA) is represented by P (22). The 4.1k (4.1-kDa) peptide is proteolytically cleaved from the *gag-pol* precursor to generate pp32. The IN-related in vitro-translated proteins are represented at the bottom. The abilities of these proteins to bind to nucleic acids are tabulated at the right. ds DNA binding represents binding to native DNA-cellulose. RNA binding represents binding to poly(U)-Sepharose. [M], NaCl concentration at which the proteins eluted from poly(U)-Sepharose; +++, binding and elution at a high NaCl concentration (binding emulating that of IN); ++, binding and elution at an intermediate NaCl concentration; +, minimal binding and elution at a low NaCl concentration; −, no significant binding; ±, elution at <0.2 M NaCl.

by using [$^3$H]Val. Sequencing revealed that [$^3$H]Val eluted in cycles 7 and 8, indicating initiation of translation at amino acid 27 with retention of the initiator Met on the polypeptide. The reason why the initiator Met was not removed, as described for p32, is unknown. However, DNA sequence analysis of the gene for IN and immunoprecipitation studies strongly suggest that the initiator Met at residue 27 was utilized. The COOH-terminal products C14.4K and C12.0K were produced by deletion of the 5′ end of the gene for IN down to the *Nru*I site (Fig. 3). They are cotranslated from the same vector, with C14.4K initiating at the first Met codon, corresponding to amino acid 155. This was confirmed by $NH_2$-terminal sequencing with [$^3$H]Leu. [$^3$H]Leu eluted in cycles 7, 8, and 15, which is consistent with initiation at amino acid 155 with loss of the initiator Met. C12.0K presumably resulted from readthrough of the first ATG and initiation from an internal Met codon at amino acid 177. DNA sequencing of the PrA gene for IN substantiated the presence of Met codons corresponding to amino acid residues 155 and 177. $NH_2$-terminal sequencing of the C12.0K peptide was unsuccessful. However, immunoprecipitation ensured that the proper reading frame was maintained (data not shown). Both COOH-terminal products, C14.4K and C12.0K, were immunoprecipitated with a peptide antiserum directed against amino acids 276 to 286. Hence, translation of both products proceeded to the COOH terminus of p32.

**DNA-cellulose affinity chromatography.** Native DNA-cellulose affinity chromatography (1) was used to determine whether the in vitro-translated IN-related products were capable of binding to DNA. DNA-cellulose affinity chromatography has been used to purify a variety of nucleic acid-binding proteins (1) and more recently to identify DNA-binding domains in the human glucocorticoid receptor (20) and the *Saccharomyces cerevisiae* MATα2 protein (23). Typically, 0.6 ml of affinity medium containing approximately 1.2 mg of DNA per ml was used. Hence, the experiments were done in DNA excess with only nanogram amounts of IN-related proteins. The translation reaction mixtures were treated with PMSF (to prevent proteolysis) and DNase-free RNase. The lysate reaction mixtures typically contained 2 to 3 μg of mRNA and several micrograms of tRNA which could compete with DNA-cellulose for binding to IN. RNase treatment was responsible for a significant quantitative increase in binding of p36 to DNA-cellulose (data not shown). Extensive clarification of the lysates prior to loading reduced nonspecific binding. The lysates were loaded and allowed to bind to the DNA-cellulose for 15 min before being washed with buffer B. Bound proteins were eluted with a linear NaCl gradient. Binding of p36 to native DNA-cellulose is shown in Fig. 4. p36 protein bound quantitatively (greater than 90% bound) and was eluted as a single peak at an NaCl concentration of 0.42 M. Both major products from translation of p36 bound to DNA-cellulose. No significant binding of either product was detected on a cellulose column control. A third band that migrated slightly faster than the two major products was also detected and eluted at the same salt concentrations. This third p36-related peptide was immunoprecipitated with p32-specific antiserum (Fig. 2). The $NH_2$ and COOH termini of the third peptide were not rigorously identified. In vitro-translated p32 generated a similar elution profile from DNA-cellulose, including quantitative binding (greater than 90%) and elution at 0.41 M NaCl (data not shown). No binding of p32 to cellulose was detected. Hence, the 4.1-kDa peptide at the COOH terminus of p36 does not significantly alter DNA binding. These results suggested that DNA-cellulose affinity chromatography could be used to identify regions of p32 involved in DNA binding.

Truncated IN polypeptides were tested for binding to native DNA-cellulose. The results are compiled in Fig. 3. As

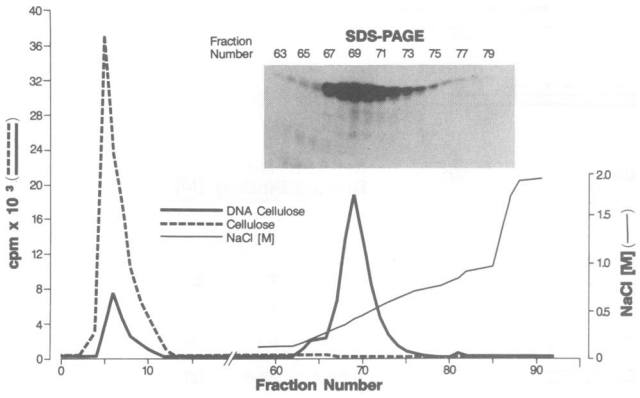


FIG. 4. Native DNA-cellulose affinity chromatography of in vitro-translated p36. Crude lysates (50 μl) of p36 were treated with RNase A and PMSF and clarified before being chromatographed on native DNA-cellulose or cellulose. Bound proteins were eluted with a 10-ml linear NaCl gradient (0.05 to 1.0 M), followed by a 2 M NaCl step wash. Fractions (0.3 ml) were collected and assayed for radioactivity in 20 μl. The counts per minute in fractions 1 to 20 represent trichloroacetic acid-precipitable radioactivity, and those in fractions 55 to 92 represent total radioactivity. (Inset) Samples (20 μl) of the fractions indicated were electrophoresed on an SDS–10% polyacrylamide gel and autoradiographed.

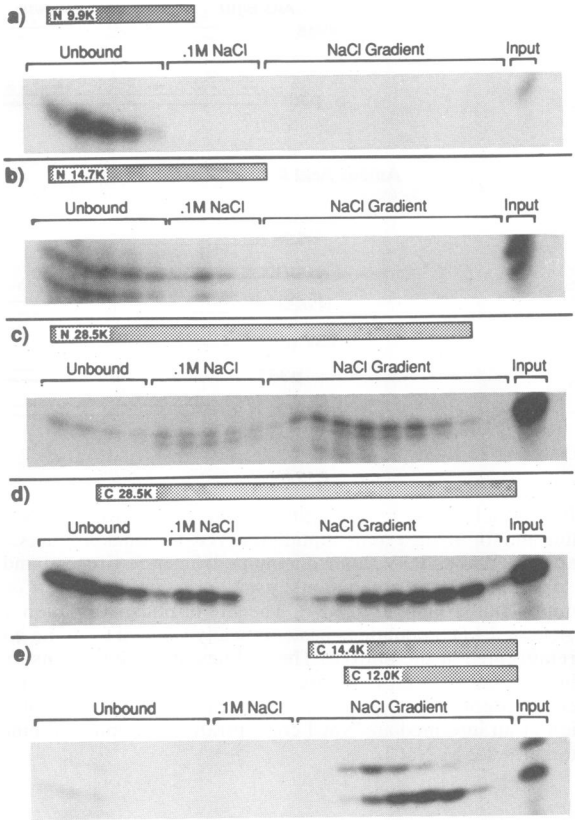


FIG. 5. Poly(U)-Sepharose affinity chromatography of truncated IN proteins. Truncated IN proteins were in vitro translated as described in Materials and Methods and chromatographed on poly(U)-Sepharose. Samples (20 μl) of 0.4-ml fractions were electrophoresed on SDS-polyacrylamide gels. Autoradiographs of SDS-polyacrylamide gels are shown. Input represents 1 to 2-μl samples of treated lysates applied to the affinity column. Stippled bars represent the truncated IN proteins assayed in the panels. Panels: a, N9.9K; b, N14.7K; c, N28.5K; d, C28.5K; e, C14.4K and C12.0K.

deletions were made from the COOH terminus of IN (N28.5K, N14.7K, and N9.9K), the ability to bind DNA was decreased. This was exemplified by a decrease in the quantity of the IN polypeptides bound to DNA-cellulose and a decrease in the NaCl concentration at which they eluted. The $NH_2$ terminus of IN contains two His residues separated from two Cys residues by 23 amino acids (Fig. 3). These His and Cys residues are highly conserved among many retroviruses, including human immunodeficiency virus 1, bovine leukemia virus, and M-MuLV (24). The homology of this sequence to the zinc fingers of transcription factor IIIA suggested that this region plays a role in DNA binding (24). However, this region appears to be insufficient for binding, as no significant binding (less than 10%) of the N9.9K product was detected on native DNA-cellulose.

The C28.5K product (Fig. 3), which presumably initiates at amino acid 27, bound quantitatively (90% bound) to native DNA-cellulose and eluted at 0.37 M NaCl (data not shown). The DNA-binding ability of C28.5K was almost indistinguishable from that of p32. However, since in vitro translation of p32 resulted in cotranslation of C28.5K, a slight difference was detected. Some of the C28.5K product eluted from native DNA-cellulose slightly ahead of p32. A similar result was obtained with p36 on DNA-cellulose (Fig. 4). Hence, deletion of the $NH_2$-terminal 26 amino acids of p32 resulted in a detectable but minor decrease in DNA-binding ability. Further, the His residues of the His-Cys box were not required for the major nucleic acid binding detected with this assay.

The C14.4K and C12.0K products did not bind significantly (≤10%) to native DNA-cellulose. As a control, all of the above-described truncated IN products were analyzed for binding to cellulose alone. No significant binding was detected, ensuring that apparent binding was a result of binding to the DNA and not to the cellulose matrix.

**Poly(U)-Sepharose affinity chromatography of IN-related proteins.** To confirm the results obtained with DNA-cellulose, the IN-related proteins were tested for binding to RNA by using poly(U)-Sepharose. Poly(U)-Sepharose affinity chromatography is a well-established procedure for purification of p32 (IN) from AMV (15) and, more recently, RSV p32 expressed in *E. coli* (42). IN has high affinity for poly(U), as indicated by quantitative binding and elution at 0.65 M KCl (15). The in vitro translation reaction mixtures were not treated with RNase A before application to poly(U)-Sepharose. Hence, competition with lysate RNA resulted in less than quantitative binding of some of the IN-related proteins. Nonetheless, p32 (IN) bound well to poly(U) and eluted at 0.60 M NaCl (data not shown; see below).

Binding of IN-related proteins to poly(U)-Sepharose is shown in Fig. 5. As with DNA-cellulose, stepwise deletion from the COOH terminus of IN resulted in a concomitant decrease in binding to poly(U) (Fig. 5a to c). The $NH_2$-terminal product, N9.9K, showed no detectable binding to poly(U). Of the two proteins shown in Fig. 5b, the slower migrating protein was identified by immunoprecipitation as the N14.7K product. The faster migrating product was presumably due to readthrough of the initiator Met and initiation at the second Met (amino acid position 27) instead. The same effect was obtained with the N28.5K product (Fig. 5c).

The elution profile of C28.5K from poly(U)-Sepharose

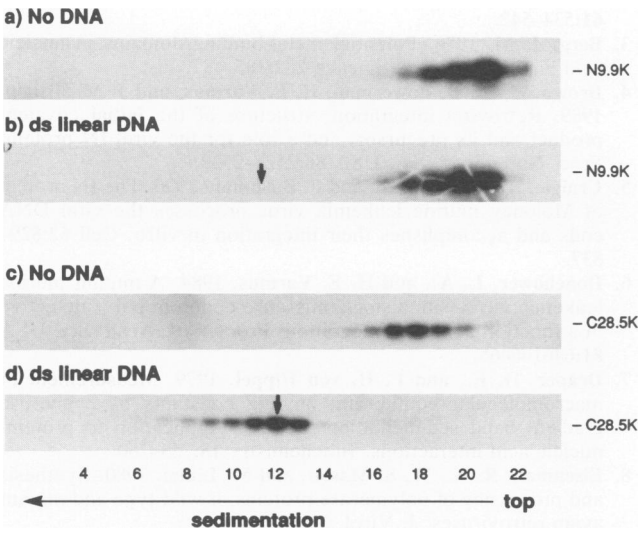


FIG. 6. Sedimentation assay of N9.9K and C28.5K products. IN-related products were incubated with or without 10 μg of linear pFT-NdeI. The reaction mixtures were sedimented on a 5 to 20% sucrose gradient. Gradient fractions were analyzed by SDS-polyacrylamide gel electrophoresis and autoradiography. The numbers refer to fraction numbers from the bottom of the gradient. The arrows indicate DNA peaks.

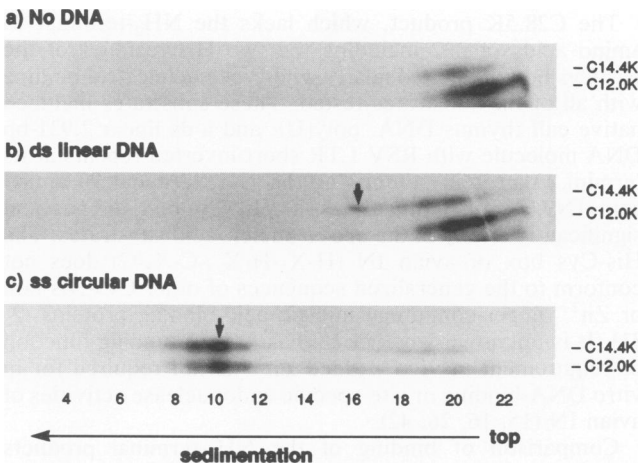


FIG. 7. Sedimentation assay of C14.4K and C12.0K products. C14.4K and C12.0K products were assayed and analyzed as described in the legend to Fig. 6. The DNA substrates used were ds linear ϕX174 and ss circular plus-strand ϕX174. The numbers refer to fraction numbers from the bottom of the gradient. The arrows indicate DNA peaks.

(Fig. 5d) was similar to that of p32. Approximately 50% of the translated product bound to poly(U) and was subsequently eluted at 0.58 M NaCl. Unlike the results obtained with DNA-cellulose, the C14.4K and C12.0K products bound quantitatively to poly(U)-Sepharose (90% bound) and resisted the 0.1 M NaCl wash. The peptides eluted at different NaCl concentrations; C14.4K eluted at 0.34 M NaCl, and C12.0K eluted at 0.43 M NaCl. Therefore, the COOH-terminal 109 amino acids of IN (C12.0K) were sufficient for binding to poly(U)-Sepharose (Fig. 5e).

**Sedimentation DNA-binding assay.** A method other than chromatography was necessary to confirm the results obtained with DNA-cellulose and poly(U)-Sepharose. The method chosen relies on sucrose gradient sedimentation (7, 40). The in vitro-translated products were incubated with an excess of a ds linear DNA substrate and subsequently sedimented on 5 to 20% sucrose gradients. Protein bound to the DNA sediments faster than free protein. The gradients were fractionated, and the fractions were analyzed by SDS-polyacrylamide gel electrophoresis (Fig. 6 and 7). The percentage of protein cosedimenting with the DNA substrate is directly related to the equilibrium binding constant (7). When the protein cosedimenting with the DNA is greater than 90% or less than 10% of the total protein, a binding constant cannot be accurately determined (7). As this was the case with the IN-related proteins, binding constants were not determined. The DNA substrate initially used in this assay (pFT-NdeI) was a 2,931-bp ds linear DNA molecule containing the RSV LTR short inverted repeats at its termini. An analogous molecule has been used as the substrate in in vitro integration reactions (9). The results of this assay using the in vitro-translated IN proteins were similar to those obtained with native DNA-cellulose chromatography. The N28.5K product, the C28.5K product (Fig. 6c and d), and full-length IN all cosedimented quantitatively (greater than 90%) with the DNA substrate. The N9.9K (Fig. 6a and b), N14.7K, C14.4K, and C12.0K products exhibited little (less than 10%) binding in this assay.

The C14.4K and C12.0K products were further analyzed with this assay by using ds linear, ss circular, and supercoiled ϕX174 DNA molecules as substrates (Fig. 7). A small percentage of the total protein cosedimented with ds linear (Fig. 7b) or supercoiled (data not shown) DNA. A large percentage (greater than 90%) of the total protein (both C14.4K and C12.0K) cosedimented with the ss circular ϕX174 substrate (Fig. 7c). These data indicate a higher affinity of the C14.4K and C12.0K products for ss over ds ϕX174 DNA molecules.

## DISCUSSION

We used in vitro translation, followed by affinity chromatography or a sedimentation assay, to identify regions of the RSV IN molecule that bind DNA. In vitro-translated p32 (IN) was indistinguishable from pp32 purified from virus. The design in cloning, immunological reactivity, and $NH_2$-terminal sequencing of in vitro-translated p32 indicate that its primary amino acid sequence was identical to that of purified RSV PrA IN. In vitro-translated IN and purified AMV IN have similar nucleic acid-binding activities, as exemplified by binding to DNA and RNA and elution profiles from poly(U)-Sepharose. Expression of proteins by in vitro translation has been used successfully to identify DNA-binding domains in a variety of proteins, including the glucocorticoid receptor, the GCN4 protein, c-Jun, and c-Fos (12, 17, 21, 32). It is not known whether in vitro-translated IN was phosphorylated or whether it possesses DNA endonuclease activity.

No significant difference between the binding of p36 and p32 to DNA-cellulose or poly(U)-Sepharose was detected. Therefore, the 4.1-kDa peptide at the COOH terminus of p36 has no apparent effect on nucleic acid binding, despite the significant net negative charge of its predicted primary amino acid sequence (38). The 4.1-kDa peptide also has no apparent effect on the endonuclease activities of bacterially expressed p32 and p36 and is not required for viral replication in tissue culture cells (25, 42).

The C28.5K product, which lacks the $NH_2$-terminal 26 amino acids of p32 including the two His residues of the His-Cys box, exhibited nearly wild-type nucleic acid binding with all of the substrates tested. These substrates included native calf thymus DNA, poly(U), and a ds linear 2,931-bp DNA molecule with RSV LTR short inverted repeats at its termini. A protein comprising the $NH_2$-terminal 90 amino acids (N9.9K), including the entire His-Cys box, did not bind significantly to any of the above nucleic acid substrates. The His-Cys box of avian IN ($H\text{-}X_3\text{-}H\text{-}X_{23}\text{-}C\text{-}X_2\text{-}C$) does not conform to the generalized sequences of other Cys-His box or $Zn^{2+}$ finger-containing nucleic acid-binding proteins (3, 31). It is unknown whether IN has a $Zn^{2+}$-binding function or requirement in vivo. Added $Zn^{2+}$ is not required for in vitro DNA-binding or site-specific endonuclease activities of avian IN (15, 16, 26, 42).

Comparison of binding of the $NH_2$-terminal products (N9.9K, N14.7K, and N28.5K) with all nucleic acid substrates showed that the COOH-terminal half of IN was necessary for efficient nucleic acid binding. Further, the COOH-terminal products were sufficient for binding to poly(U)-Sepharose or ss $\phi$X174 DNA, suggesting that this region plays a role in nucleic acid binding in vivo. The COOH-terminal products did not bind efficiently to ds DNA substrates. Therefore, upstream peptide sequences are also required for this function. The higher affinity of the COOH-terminal products for ss over ds nucleic acids suggests that the interaction is not simply an electrostatic interaction between a positively charged protein and the negatively charged phosphates of nucleic acids (Fig. 7). Further, the ability of the COOH-terminal region to bind to ss nucleic acids is consistent with the binding preferences of AMV IN and M-MuLV IN for ss nucleic acids (27, 37).

It is intriguing to consider the possible role of an ss nucleic acid-binding region of IN. During the integration event, IN appears to be responsible for the ss coupling reaction (4, 5, 10, 13) involving the viral 3′ hydroxyl terminus and the 5′ phosphoryl end of the host DNA. Also, in the avian retroviruses, the IN domain is contained within the $\alpha\beta$ reverse transcriptase molecule and is responsible for enhanced binding to RNA and DNA (11).

Of all the IN-related constructs generated in this study, none bound nucleic acids better than p32, the parent molecule. The C28.5K product exhibited nearly wild-type affinity for all nucleic acid substrates, while the other constructs had reduced affinities. Hence, the entire IN molecule is required for optimum nucleic acid interactions. Deletion of any part of the molecule may disrupt secondary and tertiary structures that may affect IN binding. The identities of these important regions and their auxiliary role in nucleic acid binding are being investigated.

## ACKNOWLEDGMENTS

We thank R. Horton for technical assistance and C. Hall for typing.

D.P.G. is supported by a grant from the National Institutes of Health (CA16312). S.R.M. was a National Institutes of Health predoctoral trainee (T32-GM-08109).

## REFERENCES

1. **Alberts, B., and G. Herrick.** 1971. DNA-cellulose chromatography. Methods Enzymol. **21**:198–217.
2. **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.
3. **Berg, J. M.** 1986. Potential metal-binding domains in nucleic acid binding proteins. Science **232**:485–487.
4. **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.
5. **Craigie, R., T. Fujiwara, and F. Bushman.** 1990. The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration in vitro. Cell **62**:829–837.
6. **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.
7. **Draper, D. E., and P. H. von Hippel.** 1979. Measurement of macromolecular equilibrium binding constants by a sucrose gradient band sedimentation method. Application to protein-nucleic acid interactions. Biochemistry **18**:753–760.
8. **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.

8a. **Fitzgerald, M., and D. P. Grandgenett.** Unpublished data.

9. **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.
10. **Fujiwara, T., and K. Mizuuchi.** 1988. Retroviral DNA integration: structure of an integration intermediate. Cell **54**:497–504.
11. **Gerard, G.** 1983. "Reverse transcriptase" in enzymes of nucleic acid synthesis and modification, vol. I, p. 1–38. CRC Press, Inc., Boca Raton, Fla.
12. **Godowski, P. J., S. Rusconi, R. Miesfeld, and K. R. Yamamoto.** 1987. Glucocorticoid receptor mutants that are constitutive activators of transcriptional enhancement. Nature (London) **325**:365–368.
13. **Grandgenett, D. P., and S. R. Mumm.** 1990. Unraveling retrovirus integration. Cell **60**:3–4.
14. **Grandgenett, D. P., T. Quinn, P. J. Hippenmeyer, and S. Oroszlan.** 1985. Structural characterization of the avian retrovirus reverse transcriptase and endonuclease domains. J. Biol. Chem. **260**:8243–8249.
15. **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.
16. **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.
17. **Halazonetis, T. D., K. Georgopoulos, M. E. Greenberg, and P. Leder.** 1988. c-Jun dimerizes with itself and with c-Fos, forming complexes of different DNA binding affinities. Cell **55**:917–924.
18. **Hewick, R. M., M. W. Hunkapiller, L. E. Hood, and W. J. Dreyer.** 1981. A gas-liquid solid phase peptide and protein sequenator. J. Biol. Chem. **256**:7990–7997.
19. **Hippenmeyer, P. J., and D. P. Grandgenett.** 1984. Requirement of the avian retrovirus pp32 DNA binding protein domain for replication. Virology **137**:358–370.
20. **Hollenberg, S. M., V. Giguere, P. Segui, and R. M. Evans.** 1987. Colocalization of DNA-binding and transcriptional activation functions in the human glucocorticoid receptor. Cell **49**:39–46.
21. **Hope, I. A., and K. Struhl.** 1986. Functional dissection of a eucaryotic transcription activator protein, GCN4 of yeast. Cell **46**:885–894.
22. **Horton, R., S. Mumm, and D. P. Grandgenett.** 1988. Avian retrovirus pp32 DNA endonuclease is phosphorylated on Ser in the carboxyl-terminal region. J. Virol. **62**:2067–2075.
23. **Johnson, A. D., and I. Herskowitz.** 1985. A repressor (MAT 2 product) and its operator control expression of a set of cell type specific genes in yeast. Cell **42**:237–247.
24. **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 enzymes. Proc. Natl. Acad. Sci. USA

**83**:7648–7652.

25. **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.
26. **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.
27. **Knaus, R. J., P. J. Hippenmeyer, T. K. Misra, D. P. Grandgenett, V. R. Muller, and W. M. Fitch.** 1984. Avian retrovirus pp32 DNA binding protein: preferential binding to the promoter region of long terminal repeat DNA. Biochemistry **23**:350–359.
28. **Krogstad, P. A., and J. J. Champoux.** 1990. Sequence-specific binding of DNA by the Moloney murine leukemia virus integrase protein. J. Virol. **64**:2796–2801.
29. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
30. **Lightfoote, M. M., J. E. Coligan, T. M. Folks, A. S. Fauci, M. A. Martin, and S. Venkatesan.** 1986. Structural characterization of reverse transcriptase and endonuclease polypeptides of the acquired immunodeficiency syndrome retrovirus. J. Virol. **60**: 771–775.
31. **Miller, J., A. D. McLachlan, and A. Klug.** 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. EMBO J. **4**:1609–1614.

31a. **Mumm, S. R., and D. P. Grandgenett.** Unpublished data.

32. **Nakabeppu, Y., K. Ryder, and D. Nathans.** 1988. DNA binding activities of three murine Jun proteins: stimulation by Fos. Cell **55**:907–915.
33. **Palmiter, R. D.** 1977. Prevention of $NH_2$-terminal acetylation of proteins synthesized in cell-free systems. J. Biol. Chem. **252**: 8781–8783.
34. **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.
35. **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.
36. **Roth, M. J., P. L. Schwartzberg, and S. P. Goff.** 1989. Structure of the termini intermediates in the integration of retroviral DNA: dependence on IN function and terminal sequence. Cell **58**:47–54.
37. **Roth, M. J., N. Tanese, and S. P. Goff.** 1988. Gene product of Moloney leukemia virus required for proviral integration is a DNA-binding protein. J. Mol. Biol. **203**:131–139.
38. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32**:853–869.
39. **Schwartzberg, P. J., 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 establishment of the integrated provirus. Cell **37**:1043–1052.
40. **Steeg, C. M., and V. M. Vogt.** 1990. RNA-binding properties of the matrix protein ($p19^{gag}$) of avian sarcoma and leukemia viruses. J. Virol. **64**:847–855.
41. **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.
42. **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.
43. **Vora, A. C., M. L. Fitzgerald, and D. P. Grandgenett.** 1990. Removal of 3′-OH-terminal nucleotides from blunt-ended long terminal repeat termini by the avian retrovirus integration protein. J. Virol. **64**:5656–5659.