JOURNAL OF VIROLOGY, Feb. 1987, p. 534–542
0022-538X/87/020534-09$02.00/0

Vol. 61, No. 2

Proteolytic Processing of Avian Sarcoma and Leukosis Viruses *pol-endo* Recombinant Proteins Reveals Another *pol* Gene Domain

FERN ALEXANDER,[1] JONATHAN LEIS,[2] DANIEL A. SOLTIS,[1] ROBERT M. CROWL,[3] WALEED DANHO,[3] MOHINDAR S. POONIAN,[3] YU-CHING E. PAN,[3] AND ANNA MARIE SKALKA[1]*

Department of Molecular Oncology, Roche Institute of Molecular Biology,[1] and Exploratory Research, Roche Research Center, Hoffman-La Roche Inc.,[3] Nutley, New Jersey 07110, and Department of Biochemistry, Case Western Reserve University, Cleveland, Ohio 44106[2]

Received 12 May 1986/Accepted 23 October 1986

Three *pol* gene products have been identified in avian retroviral particles: the full-length 95-kilodalton (kDa) β chain of reverse transcriptase and two proteolytic cleavage products of β, a 63-kDa reverse transcriptase α chain derived from the amino terminus of β and a 32-kDa (pp32) endonuclease from its carboxy terminus. By using molecularly cloned retroviral DNA and synthetic oligonucleotides to introduce initiator ATGs and codons corresponding to the authentic N termini, we constructed two bacterial-expression clones; one clone contains the entire *pol* gene, and the other contains the region encoding the pp32 domain. A 99-kDa protein was synthesized in *Escherichia coli* by the full-length clone, and a 36-kDa protein was synthesized by the endonuclease domain clone. The recombinant proteins exceeded the size of both the mature viral β chain and the pp32, respectively, by approximately 4 kDa. These larger sizes, however, are consistent with predictions from the DNA sequence of the *pol* gene. Processing of the recombinant *pol* proteins was examined by using p15 protease purified from virus particles and antisera directed against synthetic peptides corresponding to three domains in *pol*. Proteolytic digestion of the 99-kDa product with p15 produced a 63-kDa protein that comigrated on polyacrylamide gels with the α chain of reverse transcriptase and a 36-kDa fragment that comigrated with the endonuclease domain product. Further digestion of the 36-kDa protein yielded a 32-kDa protein that comigrated with viral pp32 endonuclease. Thus, we concluded that two p15-sensitive sites exist in *pol*. Cleavage at the previously identified site produces α, and cleavage at the newly discovered site removes approximately 4 kDa from the C terminus of the primary protein product. Since the 36-kDa protein was also detected in protein isolated from virus particles, it seems probable that processing at the C-terminal site is a normal step in the production of mature β and pp32 endonuclease products.

The reverse transcriptase (RT) of avian sarcoma and leukosis viruses (ASLV) is a multifunctional enzyme encoded by the *pol* gene. The predominant form of the viral enzyme is composed of a 63-kilodalton (kDa) α and a 95-kDa β polypeptide chain (17). The molecular size designations used herein differ slightly from those reported previously, but are presumably more accurate since they are based on the tabulated amino acid content predicted from the DNA sequence (32). Tryptic peptide analysis of both polypeptide chains suggests that the α chain is derived from the amino-terminal region of β by proteolytic cleavage (9, 30). The remaining C-terminal region released by this cleavage has been designated pp32 (31). The αβ protein possesses several enzymatic activities, including an RNA-dependent DNA polymerase, a DNA-dependent DNA polymerase, a DNA-RNA unwinding activity, and an RNase H activity (for a review of the enzymatic activities, see reference 35). It also contains a sequence-specific DNA endonuclease activity, a property it shares with both the ββ form of the enzyme and the pp32 protein, although the in vitro cofactor requirements (Mn^{2+} or Mg^{2+}) of the endonuclease reaction differ for each protein (10, 13, 20). The various *pol* gene products together are required for the synthesis of proviral DNA (35) and the integration of this DNA into the host cell genome (28).

The initial translation product of the *pol* gene is contained in a *gag-pol* fusion protein, Pr180. Proteolytic cleavage of the *gag-pol* precursor to release β is believed to occur either during or after virus budding and is performed by p15, the *gag*-encoded protein nearest the C terminus of Pr76 (8, 26, 27, 29, 32). In vitro, α and pp32 can be produced by p15 digestion of purified αβ dimers (26).

The predicted size of the *pol* gene in the Prague C strain of Rous sarcoma virus (RSV), as deduced from its nucleotide sequence, is 98.6 kDa (32). This is somewhat larger than the 93-kDa size determined for the β chain of the related avian myeloblastosis virus (AMV) (11). This size difference could reflect a variation in the *pol* gene coding capacity of RSV and AMV, differences in modification or folding of the mature RSV *pol* proteins, or, as suggested previously by Grandgenett et al. (11), additional proteolytic processing of both proteins.

To examine the structure and function of the various *pol* gene products, we cloned and expressed the entire ASLV *pol* gene and its endonuclease-encoding domain in bacterial hosts. The bacterial proteins were digested with p15 protease purified from RSV viral particles, and the products were analyzed with peptide antisera which are specific for each *pol* domain. The antisera were also used to identify authentic viral proteins analogous to the in vitro-generated p15 products. The results of the digestion studies reveal a new proteolytic site near the C terminus of the *pol* precursor protein. We propose that cleavage at this site can convert both a 99-kDa *pol* precursor to a 95-kDa β chain and a 36-kDa endonuclease domain precursor to pp32.

* Corresponding author.

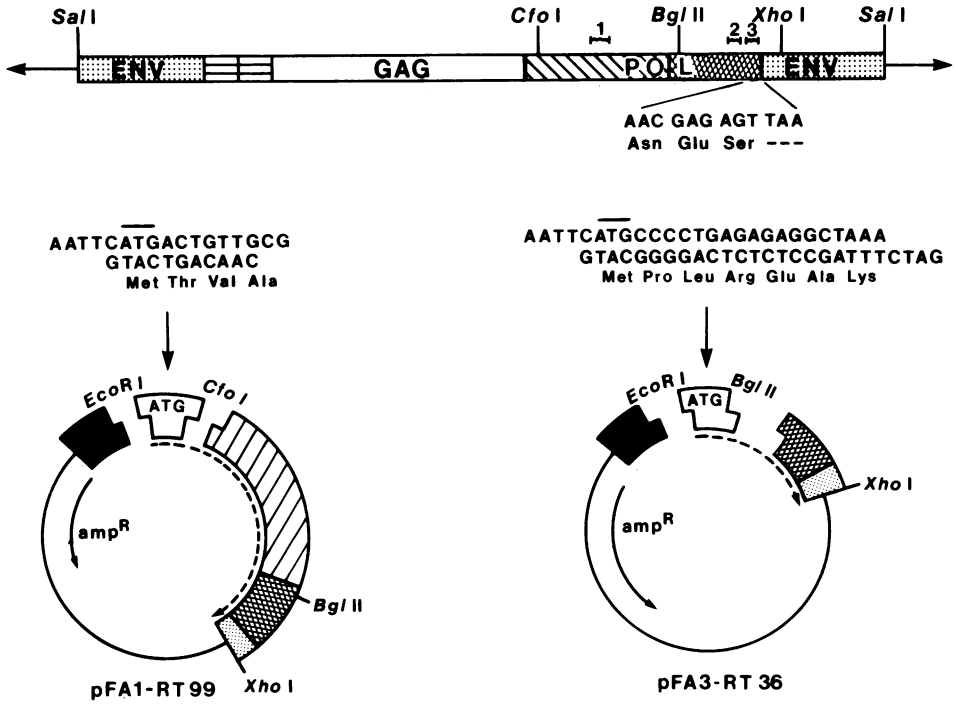

FIG. 1. Cloning strategy for the *pol* gene. The two expression clones pFA1-RT99 and pFA3-RT36 encode the full-length and endonuclease domain of the *pol* gene, respectively. *pol* gene sequences were derived from a Schmidt-Ruppin B clone (5) (map shown at the top). The locations of three peptides used to immunize rabbits are indicated by numbered short wavy lines (above the map). The last set of codons (ending with TAA) shows the natural end of the *pol* gene included in both clones. The two synthetic deoxyoligonucleotide linkers used to reconstruct the 5′ end of each gene are also shown as are the predicted amino acids in the N termini of the reconstructed genes. The initiator ATG is overlined. Symbols: ■, λ p_L promoter; ATG, synthetic linker; ▨ first two-thirds of *pol* gene; ▩, endonuclease region of *pol*; ▤, viral envelope sequences; ▭, viral *gag* sequences; ▬, LTR (long terminal repeats); ———, pBR322-derived sequences.

MATERIALS AND METHODS

Cloning of *pol* gene products for expression. Two expression clones were constructed by using a pBR322 subclone containing the entire *pol* gene from a transformation-defective mutant of the Schmidt-Ruppin B strain of RSV (6), ending at the *Xho*I site located 70 base pairs downstream of its natural terminator (Fig. 1). The *Xho*I site was ligated to the *Ava*I site of pBR322. All restriction enzymes were purchased from New England BioLabs, Inc., Beverly, Mass., and the reaction conditions suggested by the manufacturer were used.

Initiator ATGs were introduced by using synthetic deoxyoligonucleotides containing 5′-terminal *Eco*RI "sticky ends," an ATG, and codons corresponding to the normal N termini of the proteins to be expressed. Deoxyoligonucleotides were produced with a synthesizer (Applied Biosystems) according to the phosphite method (1a, 24), in which the initial 3′-terminal nucleoside was attached to a controlled pore glass support (CPG long-chain alkylamine resin; Pierce Chemical Co., Rockford, Ill.) The oligonucleotides were deblocked from the resin and purified by electrophoresis on 20% denaturing polyacrylamide-urea gels. Their purity was confirmed by electrophoresis of ^{32}P-labeled portions. For the full-length clone pFAI-RT99, the linker contained the first three predicted codons of the *pol* gene (32) and ended with a *Cfo*I-staggered 3′ end, which was ligated to the corresponding *Cfo*I site at the 5′ end of *pol*. This reconstructed the normal N terminus of α and β. Similarly, the synthetic linker constructed for the pp32 domain clone, pFA3-RT36, contained the first six predicted codons of pp32 (4) and ended in a *Bgl*II restriction site. This was ligated to the corresponding site at the 5′ end of the *endo* domain, thus reconstructing the normal N terminus of pp32. The 5′ termini of the reconstructed genes were ligated to the λ p_L promoter and a synthetic ribosome-binding site, which had a terminal *Eco*RI site, in a pBR322-based expression plasmid (pRC23) described previously (5). This vector has been shown to provide efficient bacterial expression for several eucaryotic proteins.

Recombinant plasmids were introduced into strain MC1061 (3) or KRR123, a low-protease derivative of *Escherichia coli* RRI (25); both of these contain the plasmid pRK 248*c*Its, which expresses a temperature-sensitive λ *c*I repressor. Bacteria were grown to an optical density at 550 nm of 0.4 at 30°C in M9 medium (22), which was supplemented with leucine (50 μg/ml), uridine (50 μg/ml), and thymidine (100 μg/ml) and contained ampicillin (100 μg/ml); expression of the recombinant plasmid was induced by a temperature shift to 42°C for 4 h. Cells were collected by centrifugation (JA-14 rotor, Beckman Instruments, Inc., Fullerton, Calif.) at 6,000 rpm for 25 min at 4°C, suspended in 5% of the original volume of 10 mM Tris hydrochloride (pH 7.4) containing 10% glycerol, and stored at −20°C.

Western blotting. Uninduced or induced bacteria containing the appropriate expression plasmids were lysed by suspending cells to an optical density of 2.0 at 550 nm in 0.1 ml of sample buffer (50 mM Tris hydrochloride [pH 6.8], 3% sodium dodecyl sulfate [SDS], 9% glycerol, 5% β-mercaptoethanol, 0.1% bromphenol blue). Samples were boiled for 15

min and fractionated (by using 0.2 optical density [550 nm] units of cells per lane) by electrophoresis on 7.5 to 20% gradient SDS-polyacrylamide gels, according to the procedure of Laemmli (19). Samples of virus were treated identically, and 40 μg of viral protein was applied per lane. We used purified AMV-RT (Life Sciences, Inc., St. Petersburg, Fla.) as an RT standard; the purified pp32 was a gift from D. Grandgenett. The proteins in the gels were transferred electrophoretically onto 0.2-μm-thick nitrocellulose sheets (Schleicher & Schuell, Inc., Keene, N.H.), in the presence of 12.5 mM Tris hydrochloride–96 mM glycine–0.02% SDS–20% methanol (pH 8.7), as previously described by Towbin et al. (34). The nitrocellulose filters were incubated at room temperature in blocking solution containing 5% bovine serum albumin, according to the modifications of Butnick et al. (2). They were then incubated overnight at 37°C with the appropriate antiserum (as indicated in the figure legends), at dilutions sufficient to reduce background. After extensive washing at room temperature for 2 h in phosphate-buffered saline (PBS), the filters were incubated for 1 h at room temperature with PBS containing 5% bovine serum albumin and 10^6 cpm of ^{125}I-labeled protein A or G per ml (18 to 32 mCi/mg) (Amersham Corp., Arlington Heights, Ill.). They were subsequently washed in several changes of PBS containing 0.2% Triton X-100 and dried, and autoradiograms were prepared by using preflashed XAR-5 film (Kodak).

Amino acid sequencing. Proteins were transferred electrophoretically from SDS-polyacrylamide gels to a fiber glass sheet by the method of Aebersold et al. (1). The N-terminal sequence analysis of individual bands of transferred protein was determined by Edman degradation with an automated gas-phase sequencer (model 470A; Applied Biosystems) (16). PTH amino acids were separated by high-pressure liquid chromatography on a Altex Ultrasphere ODS column, as described by Hawke et al. (15).

Purification of viral p15 protease. RSV (Prague C) was grown as previously described (33). Protease p15 was purified as part of the procedure for protein phosphatases (21). Briefly, p15 protein was solubilized from the virus (1 g) after extraction with chloroform–methanol–0.9% NaCl. The aqueous phase (120 ml) containing p15 was dialyzed against 2 liters of 10 mM β-mercaptoethanol for 4 h at 4°C, and the protein was lyophilized. Lyophilized protein was dissolved in 15-ml of 10 mM Tris hydrochloride [pH 8.0]–5 mM β-mercaptoethanol–0.1 mM EDTA–0.1% Triton X-100 and passed through a Sephadex G-75 column (5 by 36 cm) equilibrated with the same buffer. Protease p15 eluted between 53 to 60% of the column volume. Protein phosphatases contaminating the p15 preparation were removed by passing the protein through a DEAE-cellulose column (21). The p15 which did not bind to the resin was collected, dialyzed against 1 liter of 10 mM β-mercaptoethanol, and lyophilized. (Lyophilized p15 protease stored at −70°C has been stable for 2 years.) Coomassie blue staining of p15 (25 μg), which had been subjected to SDS-polyacrylamide gel electrophoresis (21), revealed only one detectable band.

p15 digestions. Bacterial lysates of recombinant clones were prepared by suspending 20 to 35 optical density (550 nm) units of induced cells in 400 μl of 10 mM Tris hydrochloride (pH 6.5) containing 5% sucrose. Cells were lysed by incubation with 200 to 400 μg of lysozyme (Cooper Biomedical, Inc., West Chester, Pa.) at 4°C for 10 min, followed by three freeze-thaw cycles, at alternating −70 and 37°C temperatures. After a 30-min centrifugation at 15,000 to 65,000 × *g*, the DNA-containing viscous pellet was discarded, and the clear supernatant was retained. The supernatants, which contained the soluble *pol* gene products, were boiled for 15 min in the presence of 1 mM dithiothreitol to inactivate bacterial proteases. They were subsequently incubated at 37°C with 50 to 150 μg of purified p15 per ml in the presence of p15 digestion buffer (50 mM Tris hydrochloride [pH 6.0], 0.25 M NaCl, 0.1% Triton X-100, 0.1 mM EDTA). Portions were removed at 60 min, 120 min, and 5 h; the reaction was stopped by adding an equal volume of ice-cold sample buffer and then boiling for 10 min. In experiments with partially purified recombinant *pol* gene products (manuscripts in preparation), the proteins were dialyzed against p15 digestion buffer and then treated as described above. Purified AMV-RT (Life Sciences, Inc.) was used directly. Samples were normalized for the amount of *pol* protein they contained relative to that of the bacterial lysates. After digestion, samples were subjected to SDS-polyacrylamide gel electrophoresis on 7.5 to 20% gradient gels.

Peptide-specific antiserum purification. Three peptides corresponding to amino acid sequences in the RSV (Prague C) *pol* gene (32) were synthesized. An additional cysteine residue not present in the viral protein was added to the N termini of peptides 2 (amino acid residues 840 to 854) and 3 (residues 879 to 890) for conjugating these peptides to keyhole limpet hemocyanin (KLH; Calbiochem-Behring, La Jolla, Calif.) by using *m*-maleimidobenzoyl-*N*-hydroxysuccinimide ester (MBS; Boehringer GmbH, Mannheim, Federal Republic of Germany). (The sequence and origin of these two peptides are shown in Fig. 7 below). Peptide 1 (corresponding to amino acid residues 276 to 290) was conjugated to thyroglobulin by using the glutaraldehyde method at a ratio of 103 μg of peptide to 1 mg of thyroglobulin. Peptides 2 and 3 were conjugated to KLH (14) at a ratio of 2 mg of peptide to 1.5 mg of KLH-MBS.

New Zealand White rabbits were immunized with the conjugated peptides by intradermal dorsal injections. On day 0, 100 μg of peptide coupled to KLH or thyroglobulin in complete Freund adjuvant was injected. The rabbits were boosted at 4-week intervals by injecting 50 μg of peptide conjugated to KLH or thyroglobulin in incomplete Freund adjuvant until a sufficiently high titer was obtained. Peptide titers were determined by a two-step enzyme-linked immunosorbent assay.

Peptide-specific antibodies were purified from the corresponding rabbit sera by passage over a protein A-Sepharose column (Pharmacia, Inc., Piscataway, N.J.). Fractions containing peptide-specific antibodies were pooled and dialyzed in PBS (20 mM sodium phosphate [pH 7.4], 100 mM NaCl). This pool was then applied to a column (Affi-Gel 10; Bio-Rad Laboratories, Richmond, Calif.) to which the appropriate peptide had been linked, in accordance with instructions of the manufacturer. Both columns were run at room temperature, and the antibodies were eluted by using 1 M acetic acid–0.15 M NaCl. The affinity-purified antibodies were then dialyzed in PBS and stored at 4°C.

RT antiserum. Polyclonal goat antiserum prepared by immunization with purified AMV-RT (765-168) was obtained from the National Cancer Institute Office of Program Resources and Logistics.

Virus purification. Supernatants of chicken embryo fibroblasts infected with the avian leukosis virus (ALV) Rous-associated virus-2 (RAV-2) were collected and stored at 4°C. After thawing, the cell debris was removed by centrifugation at 3,000 rpm for 10 min (JA-14 rotor; Beckman Instruments, Inc.). Virus was harvested by centrifugation (type 19 rotor; Beckman Instruments, Inc.) at 19,000 rpm for 2.5 h at 4°C.

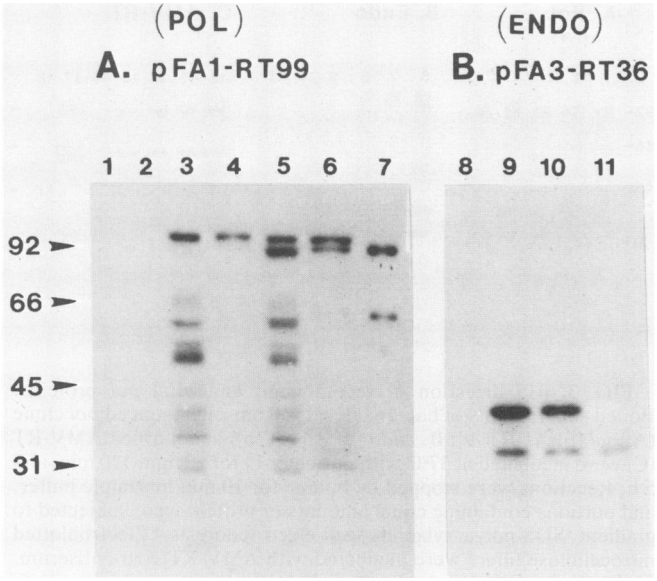

FIG. 2. Identification of the proteins expressed by clones pFA1-RT99 and pFA3-RT36. Uninduced or induced bacteria containing the appropriate plasmids were lysed and prepared as described in the text. Samples (0.2 optical density [550 nm] units of cells) were applied to an SDS-polyacrylamide gradient gel (7.5 to 20%); after electrophoresis, the gel was electroblotted onto nitrocellulose. The Western blot was incubated with goat antiserum directed against purified AMV-RT. Arrows indicate molecular weights of standard protein markers. (A) Clone pFA1-RT99. Lanes: 1, negative control (induced pRC23); 2, uninduced *pol* clone; 3, induced *pol* clone; 4, recombinant *pol* protein partially purified from inclusion bodies; 5, mixture of induced *pol* clone cell lysate and purified viral AMV-RT; 6, recombinant *pol* proteins partially purified by detergent extraction of inclusion bodies; 7, AMV-RT. (B) Clone pFA3-RT36. Lanes 8, uninduced *endo* clone; 9, induced *endo* clone; 10, partially purified *endo* protein (manuscript in preparation); 11, purified viral pp32.

Pelleted virus was suspended in TEN buffer (10 mM Tris hydrochloride [pH 7.5], 10 mM NaCl, 1 mM EDTA) and purified by sedimentation through a 15-ml, 20 to 60% linear sucrose density gradient (SW27 rotor; Beckman Instruments, Inc.) for 10 h at 25,000 rpm. The cloudy band containing the virus was diluted fivefold with TEN buffer and pelleted by centrifugation (70 Ti rotor; Beckman Industries, Inc.) at 33,000 rpm for 1 h. Virus pellets were suspended in TEN buffer and stored at −70°C.

RESULTS

Molecular cloning and expression of ASLV *pol* gene products in *Escherichia coli*. Two bacterial expression clones were constructed to provide a full- and partial-length product of the *pol* gene; one contains the entire *pol* gene, and the other contains only the pp32 encoding domain. DNA-containing segments of the ASLV *pol* gene were obtained from a previously described molecular clone of the Schmidt-Ruppin B strain of this virus (6). In virus-infected cells, the initial translation product of the *pol* gene is Pr180, a *gag-pol* fusion protein. Consequently, the *pol* gene lacks its own ATG initiation codon. Two synthetic oligonucleotides were constructed to introduce initiator ATG codons directly upstream of reconstructed N termini (Fig. 1) identical to those previously determined for viral β and pp32 (4, 32). These were joined to the remainder of the *pol* coding sequences derived from a cloned *pol* gene.

Recombinant plasmids were introduced into a low-protease derivative of *E. coli* RRI (25), which expresses a temperature-sensitive λ *c*I repressor. Recombinant proteins were expressed upon induction of transformed cultures, as determined by comparison of protein-stained gels from uninduced and induced cultures (not shown here). The pp32-encoding plasmid expresses the recombinant protein at a level of approximately 10% of the total bacterial protein. The expression of the full-length recombinant *pol* gene product is approximately fivefold less. In this clone, however, the majority of the *pol* protein is located in insoluble inclusion bodies (manuscript in preparation). For the subsequent experiments, we used either the soluble protein fraction or the full-length protein, which can be partially purified from the inclusion bodies by differential salt extraction.

pol-derived products were identified by Western blotting, using goat polyclonal antibodies prepared against purified AMV-RT (Fig. 2). As expected, no *pol*-related proteins were detected for bacterial lysates containing the pRC23 vector or for the uninduced pFAl-RT99 *pol* clone (Fig. 2, lanes 1 and 2). However, several *pol*-related proteins were produced by the induced full-length *pol* clone (lane 3). The largest and most intensely labeled product migrated as a protein of approximately 99 kDa. Smaller minor bands were also observed, some of which may have resulted from degradation of the 99-kDa protein by bacterial proteases, by premature termination, or by internal initiation of translation. One protein, migrating at the position of 66 kDa, could represent the product of an internal transcription initiation observed by others (H. Ghosh, personal communication).

The 99-kDa bacterial protein product (Fig. 2, lane 4) was the size predicted from the coding capacity of the RSV (Prague C) *pol* gene (32). Based on gel analysis, however, it exceeded the size of the mature β chain of RT purified from AMV virions by approximately 4 kDa (lane 7). This size discrepancy is not due to a gel artifact, since the 99- and 95-kDa proteins can be resolved from one another when RT is mixed with induced bacterial lysate (lane 5). The protein was solubilized by a nonionic detergent wash of inclusion body proteins (lane 6) (23), and in this instance, not only was the 99-kDa product found, but a protein which migrated with the mature viral 95-kDa β chain was also detected.

Results of a Western blot of lysates from both uninduced and induced bacteria containing the pp32 domain expression plasmid show that two *pol*-related products are produced upon induction; the major one migrated at the position of 36 kDa, the size predicted from the coding capacity of the cloned segment (Fig. 2B). This suggests that the additional 4 kDa of protein seen in the full-length *pol* product is encoded within the endonuclease domain. A minor product migrated in a manner similar to authentic viral pp32 protein (compare Fig. 2B, lanes 9, 10, and 11). The amino acid sequence of the first 12 residues in gel-purified 36-kDa protein was determined to be Pro-Leu-Arg-Glu-Ala-Lys-(Asp)-Leu-His-Thr-Ala-Leu. . . . Apparently, the initial methionine residue encoded in the recombinant construct was removed in the bacterial host so that the N terminus of the bacterially produced p36 is identical to that of viral pp32, which is produced by proteolytic cleavage of a larger precursor. This phenomenon is not unusual for eucaryotic recombinant proteins expressed in bacteria. Amino acid analysis of gel-purified 32-kDa protein gave the same N-terminal sequence. Thus, the difference in size of these two proteins must reflect an absence of amino acids from the C terminus, a result of limited proteolytic degradation, or a premature chain termination in the induced bacteria.

Recombinant ASLV *pol* products can be processed by purified viral p15 protease. Processing of the *gag* and *gag-pol* precursors is believed to take place after virus budding via proteolytic cleavage by the *gag*-encoded p15 protein (8, 26). p15 has been characterized as a thiol protease, with optimum in vitro activity at pH 5.7 (7). Recombinant *pol* proteins could be cleaved by exogenously added, purified RSV p15 protease to yield products which comigrated with mature viral α, β, and pp32 peptides (Fig. 3). Western blot analysis of bacterial lysates containing the 99-kDa recombinant protein exposed to p15 for increasing lengths of time (Fig. 3A) showed an initial product which migrated at 95 kDa (as the β chain of AMV-RT); moreover, its intensity increased with length of exposure to the protease. To eliminate confusion of p15 digestion products with those of possible endogenous bacterial proteases, all samples were boiled before addition of p15. Incubation at 37°C, without p15 protease showed a pattern identical to that of the initial samples (data not included).

Other bands were also evident after incubation of the lysate containing the 99-kDa recombinant protein with p15. One migrated at 63 kDa as expected for the α chain. Proteins which migrated as 36 kDa and 32 kDa could also be detected (Fig. 3, lanes 2 to 5). The 36-kDa species appeared transiently (lanes 2 to 4), and presumably it was then cleaved to form the 32-kDa protein (lane 5). A smaller 30-kDa species was also produced. It may have come from cleavage at a site within the 36-kDa protein, which is susceptible to p15 (or a cellular protease contaminant of the p15 preparation). The same site may be recognized by a bacterial protease, insofar as a protein of equal size was also present in bacterial lysates of the endonuclease domain clone. The 66-kDa protein, believed to have originated from an internal initiation (as noted above), also appeared to be sensitive to p15 (lane 5). Its predicted products, of approximately 30-kDa, would be expected to migrate with others in that size range. Loss of 99-kDa material coincided with the appearance of the major new bands.

The *pol-endo* recombinant product could also be digested by p15 to form a protein that comigrated with viral pp32 (Fig. 3B). After p15 digestion, most of the 36-kDa species disappeared, leaving the 32-kDa band and the smaller 30-kDa product. Since the 99-kDa and 36-kDa recombinant proteins can be cleaved by p15 protease to yield 95-kDa and 32-kDa-size species, respectively, we conclude that a proteolytic site must be located approximately 4 kDa from their identical C termini.

In results with RT purified from AMV virions that were treated with purified RSV p15 protease, little change was detected (as expected), since the RT is presumably formed by p15 processing which occurs in the virions (Fig. 3C). However, after extensive digestion of viral RT with a larger amount of p15 protease, both 36- and 32-kDa products were detected (lane 16). Lengthy exposure of products of the less exhaustive digestion (lane 17) also revealed material which comigrated with the 36-kDa recombinant protein. This suggests that small amounts of a larger *pol* precursor containing the C-terminal 4-kDa domain were present in the viral RT preparation.

Identification of recombinant and viral *pol* products by using specific peptide antisera. To confirm the origin of the recombinant *pol* p15 cleavage products, we treated samples separated on polyacrylamide gels with peptide-specific polyclonal rabbit antisera (peptide sequence origins, Fig. 1; sequences for peptides 2 and 3, Fig. 7 [see below]). The reactivity of the sera, described below, confirms predictions

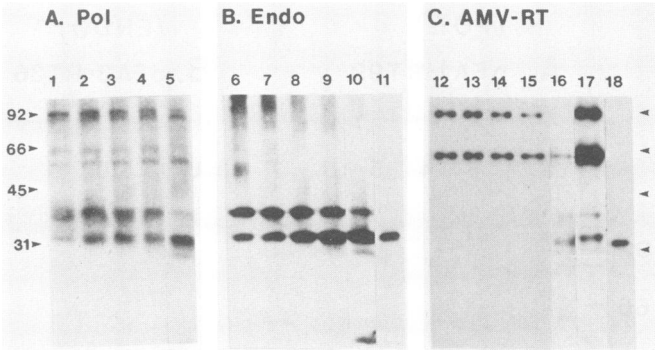

FIG. 3. p15 digestion of recombinant and viral *pol* proteins. Boiled supernatants of bacterial lysates from each induced *pol* clone (A, *pol* [pFA1-RT99]; B, *endo* [pFA3-RT36]) and purified AMV-RT (C) were incubated at 37°C with purified p15 for 60 min, 120 min, and 5 h. Reactions were stopped by boiling for 10 min in sample buffer, and portions containing equal amounts of protein were subjected to gradient SDS-polyacrylamide gel electrophoresis. Electroblotted nitrocellulose filters were incubated with AMV-RT goat antiserum. Molecular weight standards are indicated by arrows. Lanes: 1, 6, and 12, incubation at 37°C for 180 min without p15; 2, 7, and 13, incubation with p15 for 60 min; 3, 8, and 14, incubation with p15 for 120 min; 4, 9, and 15 incubation with p15 for 5 h; 5, 10, and 16 incubation with two times more p15 for 5 h; 17, long exposure of lane 15; 11 and 18, purified viral pp32.

of the amino acid sequence based on DNA sequence analysis of the Prague C strain of RSV (32). Antiserum against peptide 1 (predicted amino acids 276 to 290) detects proteins that contain the α domain of RT, and antiserum against peptide 2 (predicted amino acids 840 to 854) detects the pp32 domain. Peptide 3 corresponds to a 12-amino-acid sequence near the predicted C terminus of the 99-kDa recombinant protein (amino acids 879 to 890). Antiserum against this peptide should only detect precursor molecules and the small C-terminal cleavage product.

The α-specific peptide 1 antiserum binds to the α and β chains of purified viral RT (Fig. 4, lane 1) and to the 99-kDa recombinant protein (lanes 2 and 4) and its 63-kDa protease digestion product (lane 3). As expected, in the absence of p15 (lane 4), no 63-kDa product was detected. There were, however, additional smaller products in the lysate (lanes 2, 3, and 4), including bands at 36 and 32 kDa (unaffected by p15) which were most likely, cross-reacting *E. coli* proteins. This is consistent with the observation that this antiserum does not detect endonuclease domain proteins, including the recombinant 36-kDa protein or its 32-kDa digestion product (lanes 5, 6, and 7), or viral pp32 (lane 8).

The antiserum directed against peptide 2 detected the β chain of viral RT (Fig. 4B, lane 1), the 99-kDa recombinant protein (lanes 2 and 4), the 36-kDa recombinant protein (lanes 5 and 7), as well as the 32-kDa digestion product (lanes 3 and 6). As expected, it did not detect the 63-kDa α subunit (lanes 1 and 3), but it did identify viral pp32 (lane 8).

Peptide 3 antiserum was unable to detect any protein in which the presumed C-terminal, 4-kDa portion had been removed (Fig. 4C). Thus, it did not bind to the α or β chain of viral RT or to the pp32 protein (lanes 1 and 8), whereas it did detect the 36-kDa bands (lanes 5 and 7). The peptide 3 antiserum should also be able to detect the 99-kDa recombinant protein (lanes 2 and 4); however, the concentration of the 99-kDa protein in these lysates was quite low, as is the

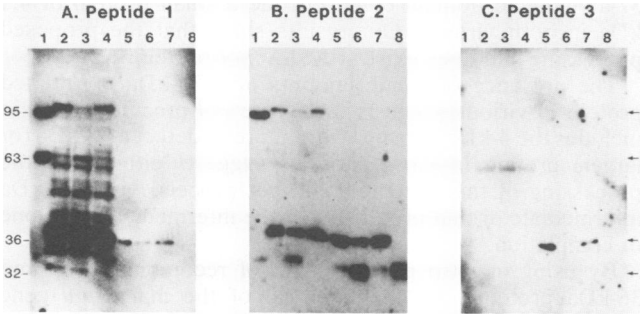

FIG. 4. Identification of p15 digestion products with peptide-specific antisera. Bacterial lysates containing protein from the recombinant clones were prepared and treated with p15 protease as described in the text, except that the lysates were not boiled before digestion and incubation was for 6 h at 37°C. The reactions were stopped by freezing (dry ice-ethanol bath); samples were then run on a 9% polyacrylamide SDS gel, and the proteins were transferred to 0.2-μm-thick nitrocellulose as described by Kyhse-Andersen (18). The filters were incubated overnight at room temperature in blocking solution (50 mM Tris hydrochloride [pH 7.5], 0.5 M sodium chloride) containing 5% bovine serum albumin and then incubated at 37°C for 1 h in blocking solution plus 3% gelatin. After being rinsed in distilled water, they were incubated for 1 to 2 h at room temperature with the indicated peptide-specific antisera. The filters were then incubated with affinity-purified, goat anti-rabbit immunoglobulin G (IgG), followed by affinity-purified, rabbit anti-goat IgG (both from Cooper Biomedical Inc.) before addition of ^{125}I-labeled protein A or G to increase the signal. After each incubation, filters were washed in PBS containing 0.05% Tween 20 (Sigma Chemical Co., St. Louis, Mo.). Results with peptide 1-specific antiserum (A), peptide 2-specific antiserum (B), and peptide 3-specific antiserum (C) are shown as follows. Lanes: 1, AMV-RT; 2, pFA1-RT99 lysate incubated (0 h) with p15; 3, pFA1-RT99 lysate incubated for 6 h with p15; 4, pFA1-RT99 lysate incubated for 6 h with no p15; 5, pFA3-RT36 incubated (0 h) with p15; 6, pFA3-RT36 lysate incubated for 6 h with p15; 7, pFA3-RT36 lysate incubated for 6 h with no p15; 8, viral pp32.

relative affinity of this antiserum. Thus, in this case, the absence of a 99-kDa band is not unexpected. In analyses with partially purified protein from the pFA1-RT99 clone, we found that peptide 3 antiserum detected only the larger of the two species identified by the AMV-RT antibody (99 kDa and 95 kDa; Fig. 2, lane 6). Thus, we concluded that the smaller of these lacks the C terminal region. Peptide 3 antiserum also detected a minor band at approximately 60 kDa, which is present even without digestion and is most likely an *E. coli* protein. The inability of peptide 3 antiserum to detect any protein which migrated at 32 kDa after p15 digestion (Fig. 4C, lane 6) confirmed that the C-terminal region of the *pol* protein was absent from the 32-kDa cleavage product and thus, presumably, from the 95-kDa product as well.

Identification of the viral *pol* products corresponding to 99 and 36 kDa. To enhance detection of *pol*-related proteins that had the *endo* domain, viral proteins were analyzed on Western blots by using a mixture of the AMV-RT polyclonal goat antiserum and our peptide 2 antiserum. The results (Fig. 5) show that with this mixture we could detect low levels of a 36-kDa protein species in ALV RAV-2 virions (lane 1), which comigrated with the 36-kDa protein expressed by the endonuclease clone (lane 4). The 36-kDa protein could also be detected in virion preparations with peptide 2 antiserum alone (data not shown). Interestingly, three other bands having sizes between 45 kDa and 65 kDa were also detected at low levels, and they comigrated with the presumed degradation products of the 99-kDa recombinant *pol* protein (lane 3). However, we were unable to detect a 99-kDa protein in the virus. We were also unable to detect a 4-kDa species in virion proteins by using peptide 3-specific antiserum; similar analysis of p15 digests of recombinant proteins (data not shown) also failed to reveal this small protein.

DISCUSSION

We have synthesized two products of the ASLV *pol* gene in bacteria; one is encoded by the entire gene and the other by its endonuclease domain. Although they are each approximately 4 kDa larger than the corresponding proteins purified from virus, their increased size is identical to that predicted by the Prague C strain nucleotide sequence. Since bacteria lack the specific viral proteases required for posttranslational processing of these proteins, the bacterial expression system provides us with an opportunity to identify the initial translation products encoded by the viral gene. The larger size of the endonuclease domain product indicates that the extra sequences reside in the C-terminal domain of the full-length product. If these reflect the sizes of the initial *pol*-encoded translation products in the virus, they provide evidence of processing intermediates which have escaped detection until now.

Cleavage of the full-length recombinant protein by viral p15 protease produced a protein that comigrated with the AMV-RT α chain on polyacrylamide gels. This demonstrates

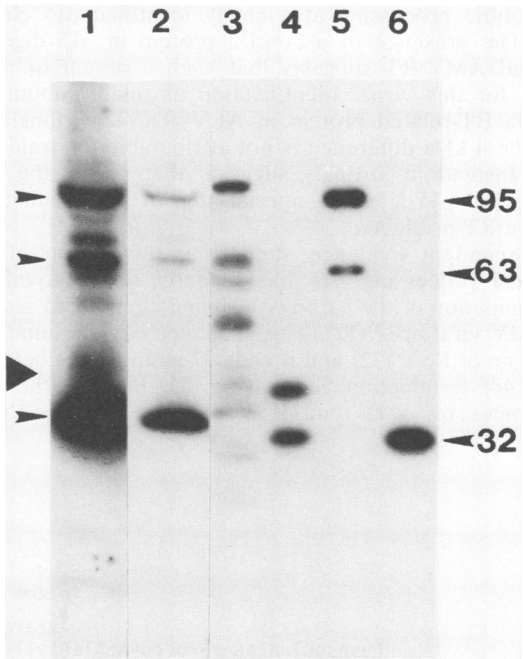

FIG. 5. Identification of viral RT-related proteins. Viral protein contained in ALV RAV-2 virions (40 μg) (lanes 1 and 2), was analyzed on Western blots, along with total bacterial lysates (0.2 optical density [550 nm] units of cells per lane) from the induced *pol* clone (lane 3) and *endo* clone (lane 4). Lane 5, AMV-RT; lane 6, viral pp32. Lane 1 is a long exposure of lane 2. The electroblotted nitrocellulose filter was incubated with a mixture of AMV-RT goat antiserum and peptide 2-specific antiserum. Arrows on the right indicate molecular weights of the viral proteins and serve as markers for lanes 3 to 6; arrows on the left indicate the location of these proteins in lanes 1 and 2. The large triangle shows the position of the 36-kDa protein.

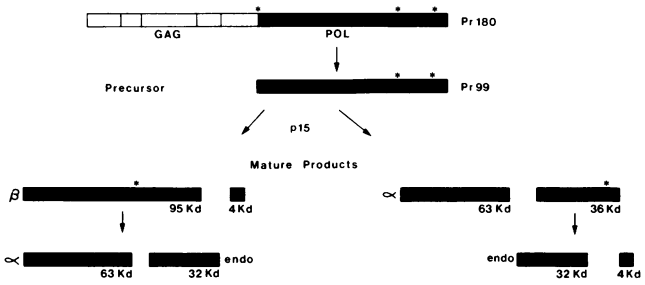

FIG. 6. Proposed schemes for processing of *pol* gene products. The diagram at the top is a map of Pr180, the initial *gag-pol* fusion protein. *pol* is cleaved from *gag* to yield a 99-kDa *pol* precursor containing two additional p15 cleavage sites (asterisks). This precursor may be digested or partially digested by p15 via two alternate pathways to generate the mature *pol* products depicted in the bottom two lines of the diagram.

that p15 is capable of processing the recombinant 99-kDa protein, in vitro, and confirmed that none of the extra sequences are located in the α domain. The two other proteolytic cleavage products noted were 36- and 32-kDa proteins which comigrated with protein produced by the endonuclease clone and viral pp32, respectively. pp32, like α, is known to contain sequences included in the β chain, and its N terminus is identical to the amino terminus of the 36-kDa recombinant protein. Thus, we conclude that the extra 4-kDa protein sequence is located at the C terminus of an RT precursor (Fig. 6) and that it is removed via proteolytic processing at a newly identified p15 cleavage site. The presence of a 36-kDa protein in p15 digests of purified AMV-RT suggests that such a precursor protein exists for this virus. Identification of small amounts of a 36-kDa RT-related protein in ALV RAV-2 virions argues that the 4-kDa difference is not attributable to strain variation. Instead, it strongly suggests that processing at the C-terminal p15 site is a normal step in the production of mature RT products.

Independent evidence supporting the notion of a C-terminal processing site has recently been provided by determination of the carboxy-terminal amino acid sequence of AMV viral pp32 (12). This sequence can be found in the *pol* gene of RSV (32) and it ends 37 amino acids before the RSV *pol* termination codon (Fig. 7). If the terminal *pol* sequences of AMV and RSV are similar, this "extra" 37-amino-acid domain could encode a small protein of 4.1 kDa in both viruses. Our results show that the proposed processing site does exist in ASLV *pol* protein.

The presence of small amounts of a 36-kDa *pol*-related protein in virions suggests that some *pol* precursor protein includes the 4-kDa domain. Our failure to detect a 99-kDa (or larger) protein in virus particles suggests either that the processing of this precursor did not proceed via a 99-kDa intermediate or that processing of the intermediate had gone to completion.

By using in vitro p15 digestion of recombinant 99- and 36-kDa proteins, we identified all of the major *pol* gene products seen in virus particles. These include proteins of 95, 63, 36, and 32 kDa. What is the pathway by which these products are processed? Figure 6 shows the different products predicted from cleavage at either one of the two p15 sites identified in *pol*. Cleavage at the first site would produce the α polypeptide chain and p36, and cleavage at the second site would produce β and the 37-amino-acid, C-terminal peptide (Fig. 6). It seems possible that an α and β formed in this way might join to produce mature RT. Further cleavage of p36 at the second site would then produce pp32 and a second molecule of the C-terminal peptide; this scheme is only hypothetical. It is also possible that precursor or β chains dimerize before α and pp32 are produced. However, our results, which show p15 processing of boiled (denatured) p99, suggest that dimerization is not required for proteolytic processing.

We have been unable to detect the 37-amino-acid, C-terminal peptide either in virions or in p15 digests of bacterial precursor p99 or p36 proteins. The peptide might be subject to further cleavage by p15 or be undetectable for some technical reason. This is being explored through other experimental approaches. A comparison of the amino acid sequences around the presumed p15 cleavage sites in *gag* and in β (4, 12, 32, 36) revealed no strict consensus among these sequences. However, they are all hydrophobic, and they frequently contain a methionine residue at the cleavage site and a proline residue in the neighboring four or five amino acids. The C-terminal 37 amino acids include two proline residues which may signal further p15 processing (Fig. 7). However, the region is not hydrophobic.

The processing site which removes the C-terminal 37 amino acids is located in a very busy region of the viral genome, occurring 1 nucleotide away from the splice acceptor site that is used to create the spliced envelope mRNA.

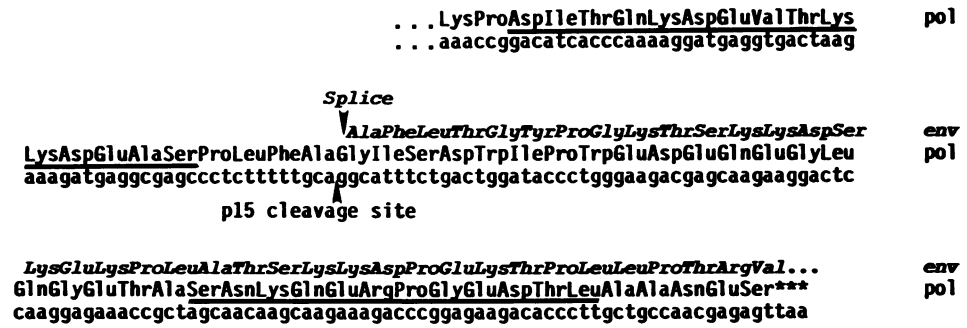

FIG. 7. C terminus of the *pol* gene. The overlapping *env* (italicized) and *pol* predicted amino acid sequences (32) are shown above the nucleotide sequence for this region. The splice site for the *env* mRNA (indicated by downward arrow above the *env* sequence) is located 1 nucleotide downstream of the last amino acid predicted for pp32 (12) (indicated by upward arrow below the nucleotide sequence). Sequences corresponding to those used for preparation of antisera 2 and 3 are underlined; the termination codon for the *pol* gene is indicated by the asterisks.

Thus, the DNA which encodes this *pol* domain also encodes sequences at the N terminus of the *env* precursor protein, a region which is upstream of the *env* signal sequence, has no known function, and is presumably removed during processing in the endoplasmic reticulum. It will be of interest to know if any essential *pol* function is encoded in this overlap domain. The amino acid sequence of this region is unique when compared with that encoded by the remainder of the *pol* protein; it contains a substantial number of hydrophilic residues, especially glutamic acid (6, 36). Our preliminary evidence indicates that the 36-kDa bacterially produced protein which has this domain contains sequence-specific endonuclease activity. Details concerning its properties and those of the other *pol*-derived recombinant proteins will be reported separately (manuscripts in preparation).

ACKNOWLEDGMENTS

We thank R. Katz for helpful suggestions and R. Chizzonite for technical expertise and guidance.

J. Leis was supported by Public Health Service grant CA38046 from the National Institutes of Health.

LITERATURE CITED

1. **Aebersold, R. H., D. B. Teplow, L. E. Hood, and S. B. H. Kent.** 1986. Electroblotting onto activated glass: high efficiency preparation of proteins from analytical SDS-polyacrylamide gels for direct sequence analysis. J. Biol. Chem. **261**:4229–4238.

1a. **Beaucage, S. L., and M. H. Caruthers.** 1981. Deoxynucleoside phosphoramidites—a new class of intermediates for deoxynucleotide synthesis. Tetrahedron Lett. **22**:1859–1862.

2. **Butnick, N. Z., C. Miyamoto, R. Chizzonite, B. R. Cullen, G. Ju, and A. M. Skalka.** 1985. Regulation of the human c-*myc* gene: 5′ Noncoding sequences do not affect translation. Mol. Cell. Biol. **5**:3009–3016.

3. **Casadaban, M. S., and S. N. Cohen.** 1980. Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. J. Mol. Biol. **138**:179–207.

4. **Copeland, T. D., D. P. Grandgenett, and S. Oroszlan.** 1980. Amino acid sequence analysis of reverse transcriptase subunits from avian myeloblastosis virus. J. Virol. **36**:115–119.

5. **Crowl, R., C. Seamans, P. Lomedico, and S. McAndrew.** 1985. Versatile expression vectors for high-level synthesis of cloned gene products in *Escherichia coli*. Gene **38**:31–38.

6. **Cullen, B. R., J. J. Kopchick, and D. W. Stacey.** 1982. Effect of intron size on splicing efficiency in retroviral transcripts. Nucleic Acids Res. **10**:6177–6190.

7. **Dittmar, K. J., and K. Moelling.** 1978. Biochemical properties of P15-associated protease in an avian RNA tumor virus. J. Virol. **28**:106–118.

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.

9. **Gibson, W., and I. M. Verma.** 1974. Studies on the reverse transcriptase of RNA tumor viruses. Structural relatedness of two subunits of avian RNA tumor viruses. Proc. Natl. Acad. Sci. USA **71**:4991–4994.

10. **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.

11. **Grandgenett, D. P., R. J. Knaus, and P. J. Hippenmeyer.** 1983. Antibodies against a synthetic peptide of the avian retrovirus pp32 protein and the β DNA polymerase subunit. Virology **130**:257–262.

12. **Grandgenett, D., 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.

13. **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.

14. **Green, N., H. Alexander, A. Olson, S. Alexander, T. M. Shinnick, J. G. Sutcliffe, and R. A. Lerner.** 1982. Immunogenic structure of the influenza virus hemagglutinin. Cell **28**:477–487.

15. **Hawke, D., P. M. Yuan, and J. E. Shively.** 1982. Microsequence analysis of peptides and proteins. II. Separation of amino acid phenylthiohydantoin derivates by high-performance liquid chromatography on octadecylsilane supports. Anal. Biochem. **120**: 302–311.

16. **Hewick, R. M., M. W. Hunkapiller, L. E. Hood, and W. J. Dryer.** 1981. A gas-liquid solid phase peptide and protein sequenator. J. Biol. Chem. **256**:7990–7997.

17. **Hizi, A., and W. K. Joklik.** 1977. RNA-dependent DNA polymerase of avian sarcoma virus B77. I. Isolation and partial characterization of the α, β2 and αβ forms of the enzyme. J. Biol. Chem. **252**:2281–2289.

18. **Kyhse-Andersen, J.** 1984. Electroblotting of multiple gels: a simple aparatus without buffer tank for rapid transfer of proteins from polyacrylamide to nitrocellulose. J. Biochem. Biophys. Methods **10**:203–209.

19. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **277**:680–685.

20. **Leis, J., G. Duyk, S. Johnson, M. Longiaru, and A. Skalka.** 1983. Mechanism of action of the endonuclease associated with the αβ and ββ forms of avian RNA tumor virus reverse transcriptase. J. Virol. **45**:727–739.

21. **Leis, J., S. Johnson, L. S. Collins, and J. A. Traugh.** 1984. Effects of phosphorylation of avian retrovirus nucleocapsid protein pp12 on binding of viral RNA. J. Biol. Chem. **259**: 7726–7732.

22. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

23. **Manne, V., E. Bekesi, and H.-F. Kung.** 1985. H-*ras* proteins exhibit GTPase activity. Point mutations that activate H-*ras* gene products result in decreased GTPase activity. Proc. Natl. Acad. Sci. USA **82**:376–380.

24. **Matteuci, M. D., and M. H. Caruthers.** 1981. Synthesis of deoxyoligonucleotides on a polymer support. J. Am. Chem. Soc. **103**:3185–3191.

25. **Miyamoto, C., R. Chizzonite, R. Crowl, K. Rupprecht, R. Kramer, M. Schaber, G. Kumar, M. Poonian, and G. Ju.** 1985. Molecular cloning and regulated expression of the human c-*myc* gene in *Escherichia coli* and *Saccharomyces cerevisiae*: comparison of the protein products. Proc. Natl. Acad. Sci. USA **82**:7232–7236.

26. **Moelling, K., A. Scott, K. E. J. Dittmar, and M. Owada.** 1980. Effect of P15-associated protease from an avian RNA tumor virus on avian virus-specific polyprotein precursors. J. Virol. **33**:680–688.

27. **Oppermann, H., J. M. Bishop, H. E. Varmus, and L. Levintow.** 1977. A joint product of the genes *gag* and *pol* of avian sarcoma virus: a possible precursor of reverse transcriptase. Cell **12**:993–1005.

28. **Panganiban, A. T., and H. M. Temin.** 1984. Circles with two tandem LTRs are precursors to integrated retrovirus DNA. Cell **36**:673–679.

29. **Rettenmier, C. W., R. E. Karess, S. M. Anderson, and H. Hanafusa.** 1979. Tryptic peptide analysis of avian oncovirus *gag* and *pol* gene products. J. Virol. **32**:102–113.

30. **Rho, H. M., D. P. Grandgenett, and M. Green.** 1975. Sequence relatedness between the subunits of avian myeloblastosis virus reverse transcriptase. J. Biol. Chem. **250**:5278–5280.

31. **Schiff, R. D., and D. P. Grandgenett.** 1978. Virus-coded 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.

32. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32**:853–869.

33. **Smith, R. E., S. Nebes, and J. Leis.** 1977. Production of large amounts of 35S RNA and complementary DNA from avian RNA tumor viruses. Anal. Biochem. **77**:226–234.
34. **Towbin, H., T. Staehelin, and J. Gordon.** 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA **76**:4350–4354.
35. **Verma, I. M.** 1977. The reverse transcriptase. Biochim. Biophys. Acta **473**:1–38.
36. **Vogt, V. M., R. B. Pepinsky, and L. E. Southard.** 1985. Primary structure of p19 species of avian sarcoma and leukemia viruses. J. Virol. **56**:31–39.