Primary Structure of p19 Species of Avian Sarcoma and Leukemia Viruses

VOLKER M. VOGT,* R. BLAKE PEPINSKY,† AND LAUREL E. SOUTHARD

Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, New York 14853

Received 8 February 1985/Accepted 22 April 1985

The internal structural proteins of avian sarcoma and leukemia viruses are derived from a precursor polypeptide that is the product of the viral *gag* gene. The N-terminal domain of the precursor gives rise to p19, a protein that interacts with the lipid envelope of the virus and that may also interact with viral RNA. The C terminus of p19 from the Prague C strain of Rous sarcoma virus was previously assigned to a tyrosine residue 175 amino acids from the N terminus. We have used metabolic labeling and carboxypeptidase digestion to show that the C terminus of p19 is actually tyrosine 155. This implies the existence of a sixth *gag* protein 22 amino acids in length and located between p19 and p10 on the *gag* precursor. The p19 species of some recombinant avian sarcoma viruses and of the defective endogenous virus derived from the *ev-1* locus migrate on sodium dodecyl sulfate-polyacrylamide gel electrophoresis as if they were about 4,000 daltons smaller than p19. We have elucidated the structure of these forms, called p19 β , by analysis of the proteins and determination of the DNA sequence of the *gag* gene from *ev-1* and *ev-2*. Esterification of carboxyl groups completely suppressed the differences in migration of p19 and p19 β . Peptide mapping showed the altered mobility to be determined by sequences in the C-terminal cyanogen bromide fragment of the proteins. We conclude from the DNA sequence that a single glutamate-lysine alteration is responsible for the altered electrophoretic mobility.

The protein p19 in avian sarcoma and leukemia viruses (ASLV) is one of the *gag* proteins that form the internal structure of virus particles. This protein may play a pivotal role in virus assembly. p19 acts as a matrix protein, being in close apposition to the lipid bilayer of the virus. This inference stems from cross-linking studies which have shown that p19, but not other *gag* proteins, can be covalently linked to phosphatidylethanolamine with bifunctional reagents specific for amino groups (20). The recent observation that p19 can be cross-linked to the integral membrane protein gp37 (8) further supports this notion. In addition, p19 is reported to bind specifically to ASLV RNA (14), suggesting a role in RNA packaging. The amino-terminal quarter of the protein appears to be involved in both of these reactions (12, 22). p19 is derived from the amino terminus of the primary translation product of the *gag* gene, the *gag* precursor Pr76 (6). The blocked N-terminal methionine residue of mature p19 and of Pr76 are identical (18). The C terminus of p19 is a tyrosine residue (1) and was previously assigned to amino acid residue 175 in the Prague C strain of Rous sarcoma virus (PrC-RSV) (24).

ASLV-related viruses can be isolated that are recombinants between an exogenous avian sarcoma virus and the endogenous Rous-associated virus-0 (RAV-0) (the product of the *ev-2* locus). Such recombinant viruses can be selected for by their transforming ability (presence of the *src* gene, from RSV) and their host cell specificity (presence of an *env* gene of subgroup E, from RAV-0). In one study, many of the recombinants were unexpectedly found to contain, instead of the p19 species of both parents, a functionally and structurally related protein, termed p19 β , that appeared to be 4,000 daltons smaller in sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) (27). The endogenous virus *ev-1* was later shown to also encode a p19 β protein (4).

To determine the nature of the difference between p19 and p19 β , we performed protein and DNA sequence analyses on examples of these proteins and their genes. From the data presented here we conclude that, in contrast to previous inferences, (i) the p19 polypeptide terminates at the tyrosine at residue 155, thus suggesting the existence of a sixth small ASLV *gag* protein, and (ii) the difference between the normal p19 species of exogenous viruses and the apparently smaller species of *ev-1* and the recombinant viruses is not in size, but rather in a single charged amino acid residue.

MATERIALS AND METHODS

Cells and viruses. Chicken embryo fibroblasts infected with PrC-RSV and a quail cell line producing the noninfectious RNA packaging mutant SE21Q1b (15) were grown in tissue culture as described previously (22). Procedures for preparation of radioactive viruses have also been described elsewhere (20). For fresh preparations of [³⁵S]cysteine-labeled virus, the following protocol was used. Confluent monolayers of cells (100-mm-diameter dishes) were washed twice with phosphate-buffered saline and then incubated for 30 min with periodic shaking in 1.5 ml of Earle balanced salt solution containing radioactive cysteine. The labeling solution was replaced with 6 ml of complete growth medium, and a single collection of supernatant was made after 8 h. Debris was removed from the collected medium by centrifugation (8,000 $\times$ g, 10 min), and virus was pelleted by further centrifugation (90,000 $\times$ g, 1 h) through a cushion of buffer (0.1 M NaCl, 1 mM EDTA, 50 mM triethanolamine hydrochloride, pH 7.5) containing 20% sucrose. Viral pellets were suspended in the same buffer plus 5% sucrose and used directly. [¹⁴C]glycine-labeled SE21Q1b $\times$ RAV-1 (a recombinant virus that is the result of a cross between RAV-1 and the defective SE21Q1b) was obtained from M. Linial. ¹²⁵I-labeled viral proteins were prepared with Na¹²⁵I and chloramine T at a specific activity of 20 μ Ci/ μ g.

PAGE. Viral proteins were separated by SDS-PAGE in a standard discontinuous buffer system as described previ-

* Corresponding author.

† Present address: BioGen, Cambridge, MA 02142.

ously (22). Main gels contained 15% acrylamide and 0.18% methylene bisacrylamide, and stacking gels contained 7.6% acrylamide and 0.21% methylene bisacrylamide. Before electrophoresis, samples were heated for 10 min at 60°C in electrophoresis sample buffer (2% SDS, 0.05 M Tris hydrochloride [pH 6.8], 12.5% glycerol, 1.5% 2-mercaptoethanol). Radioactive bands were visualized by fluorography after impregnation of the gels with sodium salicylate. For generating limited digestion peptide maps with cyanogen bromide, gel slices containing p19 or p19 β were treated with 10 mg of CNBr per ml (22), and cleavage products were analyzed by SDS-PAGE in gels containing 20% acrylamide–0.15% methylene bisacrylamide. For two-dimensional gels—SDS-PAGE followed by SDS-urea PAGE—analyses were performed exactly as described previously (21).

Aminoethylation of p19. Fresh preparations of [³⁵S]cysteine-labeled PrC-RSV and SE21Q1b (5×10^6 cpm), each in 150 μ l of virus suspension buffer, were disrupted with 0.1% SDS (1 mg of protein per ml). p19 and p12 were selectively precipitated with 4 volumes of ice-cold absolute ethanol and recovered by centrifugation at $9,000 \times g$ for 10 min in a swinging bucket rotor. Pellets were dried under vacuum and then suspended in 200 μ l of 6 M guanidine hydrochloride containing 0.2 M *N*-ethylmorpholine acetate (pH 8.1), which was boiled and purged with nitrogen before use. The total protein concentration in each sample was adjusted to 0.5 mg/ml with nonradioactive, concentrated virus. Samples were reduced with 4 mM dithiothreitol and free cysteines were aminoethylated with Aminoethyl-8 (Pierce Chemical Co., Rockford, Ill.) as described previously (25). After 5 h of reaction, the solution was adjusted to pH 5.5 with 2 N acetic acid and dialyzed against 0.005 N acetic acid. Samples were lyophilized, dissolved in electrophoresis sample buffer, and subjected to SDS-PAGE.

Carboxypeptidase digestions. Carboxypeptidase A (CPA) (treated with diisopropylfluorophosphate; catalog no. C-9762) and carboxypeptidase B (CPB) (catalog no. C-7261) were obtained from Sigma Chemical Co., St. Louis, Mo. Reaction mixtures of 50 μ l containing radioactive p19 and 0.7 μ g of CPA per ml or 1.7 μ g of CPB per ml in 0.15 M NaCl–0.05 M Tris hydrochloride (pH 7.5)–0.5 mg of bovine serum albumin per ml were incubated at 37°C. The reactions also contained approximately 0.003 to 0.03% Triton X-100, 0.03 to 0.3 mM EDTA, and 0.001 to 0.01% SDS from the p19 stock solutions. To stop the reaction, 100 μ l of 10% trichloroacetic acid was added, and the precipitated protein was removed by centrifugation for 10 min at $10,000 \times g$. A portion of the supernatant was counted for released radioactivity. A background value of trichloroacetic acid-soluble counts in the absence of added enzyme (typically, 2% for ³⁵S-labeled p19 and 4% for ¹²⁵I-labeled p19) was subtracted for each time point to generate the data presented in the figure. ¹²⁵I-labeled protein was prepared by chloramine T-catalyzed iodination of a mixture of p12 and p19 that had been prepared by chloroform-methanol precipitation (21) of avian myeloblastosis virus from leukemic chicken serum (for p19 and p23) or of SE21Q1b from the quail cell line (for p19 β and p19 α). The radioactive p19 species (specific activity, ca. 10^7 cpm/ μ g) were then visualized by autoradiography of a wet SDS-polyacrylamide gel and eluted in about 4 volumes of 0.02 M Tris hydrochloride (pH 7.5)–0.1% Triton X-100–1 mM EDTA. The [³⁵S]cysteine-labeled p19, aminoethylated or not (specific activity, ca. 10^5 cpm/ μ g), was prepared in a similar manner.

DNA sequencing. Cloned DNAs for *ev-1* (pGD27, in pBR322) and *ev-2* (pAS2, in pUC9) were obtained from A.

Skalka. The latter is actually a clone of the unintegrated viral DNA (RAV-0), but we refer to it as *ev-2* for the sake of uniformity, since its sequence should be identical to that of *ev-2* in the *gag* region. The plasmid DNA was purified on cesium chloride-ethidium bromide equilibrium gradients, and the desired *gag*-containing *Bam*HI fragments (for *ev-1*, *Bam*HI site in flanking DNA to nucleotide 532 and nucleotides 533 to 1917; for *ev-2*, nucleotides 533 to 1917) were isolated by electrophoresis in a low-melting agarose gel. Each fragment was cut with one of the enzymes described below and end labeled with reverse transcriptase and the appropriate ³²P-deoxyribonucleotide triphosphate or was end labeled directly at the *Bam*HI site. After being further cut with *Nru*I (nucleotide 1354), the radioactive fragments were repurified by electrophoresis on polyacrylamide gels. Sequence determinations were done by the chemical cleavage method of Maxam and Gilbert (16), with reagents giving A+G, G, C+T, and C ladders on gels 80 cm in length. Sequence ladders from the following sites were used to generate the final sequence data: for *ev-1*, *Bam*HI-533, *Xho*I-631, *Nar*I-798, and *Sau*3A-1051; for *ev-2*, *Bam*HI-533, *Xho*I-631, and *Nar*I-798. (The numbers refer to the conventional nucleotide sequence of PrC-RSV [24]. For the *Bam*HI labeling of *ev-2* and the *Sau*3A labeling of *ev-1*, sequence ladders were read in only one direction; for the other sites, ladders were generated in both directions. The sequence of *ev-2* up to the *Bam*HI site at nucleotide 532 was previously determined by R. Malavarca, R. Katz, and A. Skalka (personal communication); portions of this region for *ev-1* were also analyzed by these workers and agree with our sequence.

RESULTS

C terminus of p19. The C terminus of p19^{gag} of the Prague B strain of RSV is a tyrosine residue (1). According to the DNA sequence of the *gag* gene of PrC-RSV and the SRA strain of RSV (SRA-RSV) (24, 30), the only tyrosines that occur in the first 200 amino acid residues are at positions 15, 46, 155, 175, and 181 (according to the conventional numbering of PrC). These and other important features of the predicted amino acid sequences of this part of the *gag* gene are shown in Fig. 1. The residue at position 181 is known to be the fourth amino acid from the N terminus of p10 (11). The tyrosine at position 155 or 175 must therefore mark the end of p19, the molecular weight of which is about 17,000 (33). The DNA sequence further predicts that the only histidine before the beginning of p10 is residue 156. If the C terminus of p19 is residue 175, therefore, radioactive histidine should label the polypeptide, whereas if its C terminus is residue 155, it should not.

To decide between these two possibilities, we cultured PrC-RSV-infected cells with uniformly labeled [¹⁴C]histidine or with other radioactive amino acids as controls. Figure 2 shows the results of one such experiment in the form of a fluorogram of an SDS-polyacrylamide gel in which the proteins of purified RSV were electrophoresed. Lanes a, b, c, and d correspond to labelings with [³⁵S]methionine, [¹⁴C]histidine, [³⁵S]cysteine, and [¹⁴C]glycine, respectively. The bands marked p27, p19, p12, and p15 are the four major *gag* proteins of ASLV; a fifth *gag* protein, p10, is not readily visualized because it comigrates with p15 in PrC-RSV (21). It is immediately apparent from lane b that histidine is absent from p19, since no band is seen. By comparison, p27, which is present in virions in about the same number of copies as p19 and has only two histidine residues (24), forms a dark band on the film. Similar experiments with avian mye-

FIG. 1. Schematic representation of the ASLV *gag* gene. The top line represents the genomic viral RNA from the 5' end to the end of the *gag* gene. Numbers above the line refer to position from the end of the RNA, according to the conventional numbering of PrC-RSV RNA (24). Numbers below the line refer to the N- and C-terminal amino acid residues of the mature *gag* proteins p19, p2, p10, p27, p12, and p15, which are proteolytically cleaved from the *gag* precursor Pr76. L, Untranslated leader sequence. The arrows denote known N or C termini; the absence of an arrow (p2 and p10) indicates that the terminus has not been identified. The p19 region of the *gag* gene is shown expanded on the lower line. Amino acid residues important for this study are marked at the residue number: cysteine (C), methionine (M), tyrosine (Y), and histidine (H). For completeness, the amino acids between residues 174 and 181 are shown as well: proline (P), valine (V), glycine (G), serine (S), and leucine (L). All of the cysteine, methionine, tyrosine, and histidine residues in the 181-amino acid stretch of polypeptide are shown, except for cysteines at positions 163 and 165 in p2. The cyanogen bromide fragments described previously (22) are numbered in the boxes on the bottom line. Solid lines signify complete cleavage fragments visible by electrophoresis in SDS-urea PAGE; dotted lines signify fragments that are detectable only as partials.

loblastosis virus, as well as labelings with [^3H]histidine, yielded identical results (data not shown).

The conclusion that p19 ends at tyrosine 155 is further supported by an argument based on the distribution of

[^{35}S]cysteine label shown in lane c. A minor p19-related virion protein, p23, consists of the p19 polypeptide plus an additional stretch of amino acids at its C terminus (22). If p19 ended at residue 175, both p19 and p23 would contain seven cysteine residues, since the next cysteine after position 165 occurs far downstream, in the middle of p27. By contrast, if p19 ends at residue 155, it would contain four cysteine residues, and p23 would contain seven, since the three cysteines that lie in the 10 amino acids following 155, at residues 157, 163, and 165, would be in p23. We have repeatedly found that labeling with [^{35}S]cysteine enhances the p23 band relative to the p19 band, as seen, for example, in the experiment shown in Fig. 2 by comparison of lanes a and c. By similar reasoning, p23 should become labeled with radioactive histidine. We have observed such labeling in longer exposures of the fluorogram in Fig. 2 as well as in other experiments.

The previous assignment of tyrosine 175 as the p19 terminus (24, 30) was based on the observation that CPA removes tyrosine but no other amino acids from p19 (1). CPA is unable to release C-terminal proline, and a proline residue immediately precedes tyrosine 175 but not tyrosine 155. Secondary structure or disulfide bonding, however, could provide an alternative explanation of this observation. A cysteine occurs at position 154, and if this residue is linked to one of the other three cysteines by a disulfide bridge, hydrolysis by CPA might also be prevented. We designed the following experiment to test directly for a cysteine close

FIG. 2. Differences in electrophoretic profiles of ASLV proteins due to amino acid compositions and esterification. Proteins from PrC-RSV and from SE21Q1b $\times$ RAV-1 were dissolved in electrophoresis sample buffer and subjected to SDS-PAGE. Radioactive bands were visualized by fluorography. Lane a, [^{35}S]methionine-

labeled PrC; lane b, [^{14}C]histidine-labeled PrC; lane c, [^{35}S]cysteine-labeled PrC; lanes d-f, [^{14}C]glycine-labeled PrC; lanes g and h, [^{14}C]glycine-labeled SE21Q1b $\times$ RAV-1. Samples in lanes f and h were treated with acidic methanol before gel analysis (21).

FIG. 3. Carboxypeptidase digestion of p19. Aminoethylated, [^{35}S]cysteine-labeled p19 or p19 prepared identically but not aminoethylated was digested at 37°C with a mixture of CPA ($0.7\ \mu\text{g/ml}$) plus CPB ($1.7\ \mu\text{g/ml}$) or with CPA alone, and the release of trichloroacetic acid-soluble radioactivity was measured. The input into each reaction mixture was 500 cpm of ^{35}S -labeled p19. One hundred percent release is defined as one-fourth of the input counts, since p19 contains three internal cysteine residues in addition to the one cysteine near the C terminus. Iodinated, nonaminoethylated p19 was digested with CPA ($0.7\ \mu\text{g/ml}$) alone to show the release of the C-terminal tyrosine residue. At a concentration of CPA that was 15-fold lower than that shown here, one-half of the same plateau level of ^{125}I label was released in about 15 min. The plateau level of ^{125}I label corresponds of 52% of the input label. There are two other tyrosine residues in p19, which thus are presumably less readily iodinated than the C-terminal residue in the chloramine T-catalyzed reaction. Symbols: $\circ$, aminoethylated, ^{35}S -labeled p19 digested with CPA plus CPB; $\times$, ^{35}S -labeled p19 digested with CPA plus CPB; $\bullet$, ^{125}I -labeled p19 digested with CPA alone; $\blacksquare$, aminoethylated, ^{35}S -labeled p19 digested with CPA alone.

to the C-terminal end of p19. Freshly prepared [^{35}S]cysteine-labeled virus was dissolved in 6 M guanidine hydrochloride in the presence of dithiothreitol, and free sulfhydryl groups were then aminoethylated, a procedure that converts cysteine into a lysine analog. We reasoned that if p19 ends at residue 155, then after removal of tyrosine by CPA the resulting modified cysteine residue would be a substrate for CPB, which is specific for lysine and arginine residues.

Figure 3 shows the results of an experiment in which treatment of aminoethylated p19 with a mixture of CPA and CPB released [^{35}S]cysteine. No radioactivity was released by CPA alone, and no radioactivity was released by the mixture of enzymes when p19 that had not been aminoethylated was used as the substrate. As a further control, after the carboxypeptidase digestions for different times, trichloroacetic acid-precipitated p19 labeled with ^{35}S or with ^{125}I was analyzed by SDS gel electrophoresis and fluorography. The radioactive bands showed no evidence of internal breakage, migrating indistinguishably before and after the carboxypeptidase digestions (data not shown). In several experiments with two different preparations of radioactive p19, the maximal amount of released aminoethylated [^{35}S]cysteine was 11 to 15% of the total counts per minute in

the protein, or about half of the expected yield of 25%. The amount of CPA used was sufficient to release a plateau level of [^{125}I]tyrosine (Fig. 3), corresponding to 50% of total ^{125}I label in p19, in less than 10 min. We assume that this plateau represents all of the C-terminal tyrosine. (The efficiency of iodination of each of the three tyrosine residues in p19 is unknown.) We therefore infer that the aminoethylation was not complete, perhaps because complete reduction of the putative disulfide bonds is difficult to achieve.

In summary, we conclude from the carboxypeptidase and the metabolic labeling experiments that the C terminus of p19 is tyrosine 155. Since the next known *gag* protein begins with the serine at residue 178 (11), these conclusions imply the existence of a stretch of 22 amino acid residues, which we denote provisionally as $p2^{\text{gag}}$, that is not contained in the five major *gag* proteins. This stretch of polypeptide is relatively hydrophobic and very rich in cysteine. We presume that p23 differs from p19 by the presence of the p2 moiety at its C terminus.

Nature of the electrophoretic differences in p19 species. Several recombinant viruses derived from crosses between an exogenous transforming virus and the endogenous virus that is the product of the *ev-2* locus, RAV-0, contain instead of p19 a related species, called p19 β , that in SDS gels migrates faster than the p19 of either parent (27). The p19 of the noninfectious endogenous virus from the *ev-1* locus also migrates at this rapid rate (4). Both p19 and p19 β have the same N-terminal tryptic peptide (27). The kinetics and extent of release of [^{125}I]tyrosine from these proteins by CPA treatment were identical (data not shown), suggesting that their C termini are also the same. In one model to explain the mobility differences, internal deletions in the p19 part of the *gag* gene were hypothesized to have occurred during the multiple crossover events that must have given rise to the virus recombinants (27).

In a previous study on the *gag* protein p10, we observed by SDS-PAGE that large differences in the electrophoretic mobility of p10 from different ASLV can be completely suppressed by esterifying free carboxyl groups in acid methanol (21). This treatment did not result in any significant size change based on gel filtration. Esterification also caused the two different p19 species, in this case from PrC-RSV and SE21Q1b $\times$ RAV-1, to migrate together (Fig. 2): the mobility of p19 β was unaltered (compare lanes g and h), but the mobility of p19 was increased dramatically (compare lanes e and f). The *gag* proteins p27, p15, and p12 are hardly affected by the treatment. The *gag* protein p10, which in both viruses was obscured by p15 and p12, appeared as a more rapidly migrating band after esterification. Similar effects were obtained with p19 from avian myeloblastosis virus and p19 β from PrE21-21 and SE21Q1b (data not shown). The increased mobility of p19 after esterification suggests that this polypeptide contains one or more peculiar carboxylate residues that are not found in p19 β and that retard its mobility. It is well documented that the addition of negative charges to proteins can decrease their mobility in SDS-PAGE (32).

An independent size determination, migration in SDS-urea PAGE, further supports the conclusion that p19 and p19 β are the same size. This result is shown in Fig. 4, which presents the results of two-dimensional electrophoresis of a mixture of PrC-RSV proteins and *ev-1* proteins. The proteins were first electrophoresed in a standard SDS gel, and the gel strip was then placed on an SDS-urea gel for the second electrophoresis. The arrows point to the two p19 species which can be seen to have separated well in the first

dimension but to have moved nearly identically in the second dimension.

What portion of the p19 polypeptide is responsible for the anomalous migration of p19 relative to p19 β ? We applied the cyanogen bromide mapping technique previously described (19, 22) to answer this question. Figure 5 presents a comparison of the CNBr cleavage profiles of p19 (lane a) and p19 β (lane b). Bands containing these proteins were excised from an SDS gel and treated with CNBr under conditions that give a mixture of complete and partial cleavage products, and the resulting peptides were electrophoresed on a second SDS-polyacrylamide gel. The numbers at the left of the figure indicate the compositions of the peptides in the bands, 1 being the N-terminal and 4 the C-terminal complete CNBr fragment (see Fig. 1). This assignment of bands was derived from our previous analysis of CNBr-cleaved p19 on SDS-urea PAGE, coupled with a two-dimensional comparison of the cleavage fragments on SDS-PAGE and SDS-urea PAGE (data not shown). The only major difference in the profiles is the mobility of the fragment 4-containing partials 3-4 and 1-2-3-4, the latter being the uncleaved p19 species itself. The other fragment 4-containing partial, 2-3-4, was not resolved as a distinct band in this analysis. Bands containing only combinations of fragments 1, 2, or 3 are similar for p19 and p19 β . We therefore infer that the alteration that distinguishes

FIG. 4. Two-dimensional electrophoretic analysis of p19-related proteins. Samples containing a mixture of [3 H]tryptophan-labeled proteins from *ev-1* and from PrC-RSV were dissolved in electrophoresis sample buffer and subjected to SDS-PAGE. The gel profile at the left of the panel is an autoradiograph of a lane from such a gel. The remainder of the panel is a two-dimensional analysis of the same preparation in which proteins from a parallel gel strip were then electrophoresed in an orthogonal direction into an SDS-urea polyacrylamide gel and visualized by fluorography. Arrows denote the positions of p19 and p19 β . The proteins are identified at the left of the panel. The p10 spot visible in the two-dimensional gel is derived from *ev-1*; in this preparation, the p10 of PrC-RSV was split into two spots which were too light to be visible when photographed.

FIG. 5. Limited digestion of p19 and p19 β with cyanogen bromide. [14 C]glycine-labeled PrC-RSV and [14 C]glycine-labeled SE21Q1b were disrupted in electrophoresis sample buffer and analyzed by SDS-PAGE. Gel slices containing PrC-p19 and SE21Q1b-p19 β were excised with a razor blade and subjected to limited digestion with CNBr. Cleavage fragments were electrophoresed into 20% SDS-polyacrylamide gels and then visualized by fluorography. Lane a, PrC-p19; lane b, SE21Q1b-p19 β . Numbers at the left of the profiles denote the fragmentary compositions of cleavage products based on the nomenclature described in the legend to Fig. 1. The dotted lines mark the homologous fragments from the two viruses that migrated differently.

the two p19 forms lies within fragment 4, between the methionine predicted from the DNA sequence at residue 139 and the C-terminal tyrosine at residue 155.

DNA sequence of the p19 region in *ev-1* and *ev-2*. To identify exactly the origin of the electrophoretic difference between p19 and p19 β , we sequenced portions of the cloned DNA for *ev-1*, which has p19 β , and as a control, the same region for *ev-2*, which has a normal p19. The nucleotide sequences and predicted amino acid sequences for these viruses and, for comparison, for other known ASLV DNAs are shown in Fig. 6 and 7. Only 7 of the 465 nucleotides that code for p19 differ between PrC-RSV and *ev-1*. Of these, four are in triplets that code for different amino acids. Two of the different amino acids are chemically similar in the two viruses (Ile $\rightarrow$ Val and Val $\rightarrow$ Ala). Both of the remaining changes (Glu $\rightarrow$ Lys at nucleotide 806 and Thr $\rightarrow$ Met at nucleotide 810) occur in the C-terminal CNBr fragment of p19. Because of the known effect of adding negative charges to polypeptides (31), we propose that the Glu $\rightarrow$ Lys change is responsible for the mobility difference between p19 and p19 β . Consistent with this inference, the sequence of *ev-2* shows a glutamate at this position, as do the published sequences for other known p19-related proteins, from SRA-RSV (30), Fujinami sarcoma virus (28), myelocytomatosis virus MC29 (23), and Y73 avian sarcoma virus (13). SRA-RSV and the helper viruses for Fujinami sarcoma virus, MC-29, and Y73 all have p19 species that migrate like the PrC-RSV p19.

FIG. 6. Nucleotide sequence of the N-terminal portion of the *gag* gene of *ev-1* and *ev-2*. The top line shows the sequence determined for *ev-1*, starting from the beginning of the *gag* gene, at nucleotide 380 in the conventional numbering of PrC-RSV (24). The second line shows the sequence for *ev-2*, and subsequent lines show the sequences for other ASLV, where a dash indicates identity and an asterisk indicates no nucleotide. At the position marked by an arrow, after nucleotide 881, a sequence of six extra nucleotides occurs in SRA-RSV (CCACGC) and Y73 (CCACAG). The *ev-2* sequence from nucleotides 532 to 1100 was determined by us; the *ev-2* sequence from nucleotides 380 to 532, which is not shown, was determined by Malavarca et al. (personal communication) and is identical to that of *ev-1*. The sequence for SRA-RSV is taken from reference 30, and the sequences for the *gag* portions of the *gag-onc* fusion proteins of Fujinami sarcoma virus, myelocytomatosis virus MC29, and Y73 avian sarcoma virus are taken from references 28, 23, and 13, respectively.

FIG. 7. Predicted amino acid sequence of the N-terminal portion of the *gag* gene of *ev-1* and *ev-2*. The top line shows the predicted sequence for *ev-1*, starting at the first amino acid of p19. The second line shows the sequence of *ev-2*, and subsequent lines show sequences for other ASLV. A dash indicates identity. At the position marked by an arrow, after 167, two extra amino acids (Thr-Ala) occur in SRA-RSV and Y73. The extensive changes in the predicted sequence of p10 of SRA-RSV are due to an apparent reading frame change at nucleotide 980, which has not been independently confirmed.

In conclusion, the DNA sequence analysis, together with the cyanogen bromide peptide mapping and the effect of esterification, provide strong evidence that the glutamate at residue 143 somehow acts to retard the p19 species in SDS-PAGE.

DISCUSSION

We have presented data to show that the C terminus of ASLV p19 is the tyrosine at residue 155 on the *gag* precursor polypeptide, a conclusion that derives from both C-terminal analysis with carboxypeptidases and metabolic labeling studies. The previous assignment of the carboxy terminus to tyrosine 175 (24), which has been copied onto other published DNA sequences of *gag*-related transforming proteins (13, 23), was based on the observation that CPA quantitatively releases tyrosine but no other amino acid from p19 (1). We presume that it is not a penultimate proline residue but rather secondary structure or disulfide bonding of cysteine 154 that explains the ability of CPA to release only a single amino acid. In previous amino acid analyses of nonradioactive ASLV *gag* proteins, p19 was reported to contain 0.8 residues of histidine (avian myeloblastosis virus and PrC-RSV [10]), 0.2 (avian myeloblastosis virus), or 0.6 (MC29 myelocytomatosis virus [7]) residues of histidine, or 1.4

residues of histidine (Prague B strain of RSV [1]). We do not know the origin of these discrepancies with our own data, which imply that there are no histidines in p19.

Our result provides explanations for several observations on p19 and p19-related proteins that were less readily accounted for by the previous assignment of the C terminus to residue 175. First, the molecular weight of p19 is close to 17,000 by gel filtration in 6 M guanidine hydrochloride (33), exactly as predicted by a size of 155 amino acids. Second, the first amino acid of the *gag* protein p10 is residue 178 (11). Previously, it was necessary to invoke two neighboring cleavage sites, or an exopeptidase activity, to account for the N terminus of p10 being three amino acid residues beyond tyrosine 175. Now we assume that a single cleavage event generates the N terminus of p10 and the C terminus of the predicted p2 *gag* protein. Third, at least two classes of p19-related polypeptides that are longer than p19 at the C terminus have been observed in infected cells or in virions and in vitro after cleavage of Pr76 with the p15 *gag* protease (34). Pr32 is a transient intermediate cleavage product in the proteolytic processing scheme (6). p23 (and its equivalent, p19 α , from *ev-1* and some recombinant viruses) is a protein found in minor amounts in virions. Previously, models for the structures of these proteins required at least some

nonconservative processing, i.e., cleavages internal to the major mature *gag* proteins. We had therefore assumed, and tried unsuccessfully to demonstrate, that p23 includes all or a portion of the p10 moiety of the *gag* precursor and that Pr32 includes p10 or a portion of p27 or both. In view of the reassignment of the p19 C terminus, it now seems likely that p23 simply comprises p19 plus p2 and that Pr32 comprises p19 plus p2 plus p10. The dearth of purified p23, however, has prevented careful C-terminal analyses for all amino acids. Fourth, the previously predicted cleavage site between residues 175 and 176 did not fit the consensus sequence for other ASLV *gag* cleavages in which an amino acid with a hydrophobic side chain invariably occurs as the penultimate residue from the C terminus of the cleaved product. The cleavage site between residues 155 and 156 fits this consensus.

We have not detected in mature virus particles the stretch of amino acids that separates p19 and p10 in the *gag* precursor. Using [³⁵S]cysteine to selectively label this p2 region, however, we have purified from virions and identified by amino acid sequencing two peptide cleavage products derived from it (R. B. Pepinsky, R. J. Mattaliano, and V. M. Vogt, unpublished observations). These peptides account for 30 to 50% of the theoretical yield based on the relative abundance of the other mature *gag* cleavage products. Although its function is unknown, p2 is noteworthy because it contains near its C terminus the highly conserved sequence Pro-Pro-Pro-Tyr, which is found in approximately this location in the *gag* precursor in mammalian retroviruses bearing no sequence homology to ASLV, like murine leukemia virus (29), feline leukemia virus (9), and in the distant avian relative of these viruses, spleen necrosis virus (17), as well as in the apparently unrelated human T-cell leukemia virus (26). In murine leukemia virus (32) and feline leukemia virus, this sequence occurs in p12, and so one can speculate that p2 shares some of the properties of that protein.

We have shown that the exchange of a single amino acid residue, glutamate for lysine at position 143, drastically alters the mobility of p19 in SDS-PAGE, changing the apparent molecular weight by about 4,000. Although charge is known to affect mobility in SDS-PAGE (31), such a large effect is highly unusual. Since SDS-PAGE is used routinely and often uncritically to estimate molecular weights, our finding that the migration anomalies of p19 as well as of p10 (21) can be suppressed by esterification is of technical importance. We do not know what is special about this p19 glutamate residue. There are many other glutamates in p19, and other *gag* proteins have a similar mole fraction of glutamate. In RSV, this residue occurs in the dipeptide Glu-Glu, but this is not a unique feature; the DNA sequence predicts two other Glu-Glu dipeptides in p19, and these apparently do not contribute to the aberrant mobility, since esterification of carboxyl groups in the p19 β of *ev-1* does not alter its mobility. We presume that some feature of the p19 secondary structure that is maintained in SDS is altered by the glu $\rightarrow$ lys change at residue 143.

The conclusion that p19 and p19 β differ by a missense mutation obviates the need for the awkward postulate that deletions in the *gag* gene occur frequently in crosses of RAV-0 and RSV. Since neither parent has the p19 β protein, however, the question remains as to how this apparent alteration in amino acid residue 143 occurred. We suggest as a model to explain these results that the RAV-0 parent in this cross consisted of a mixture of virus particles containing RNA derived not only from the *ev-2* locus (RAV-0) but also from the *ev-1* locus. This locus is normally transcriptionally

inactive but can become activated spontaneously in some embryos and also can be turned on by exposure of cells to azacytidine (4). Although the virus particles formed from *ev-1* are biologically defective (4), transcription rates from activated *ev-1* and *ev-2* can be comparable (5). The biological inactivity of the *ev-1* virus and its close similarity to *ev-2* would make the presence of *ev-1* in a mixture difficult to detect in the absence of RNA fingerprinting (3). The *ev-1* genetic information in cells has been shown previously to readily recombine with that of other endogenous viruses being replicated in the same cells (3). If there was no inadvertent selection for *ev-1* sequences in the original cross, then the model proposed would predict that the stock of RAV-0 used in the cross should give rise to some p19 β molecules when infected into appropriate chicken cells.

Comparisons of the DNA sequences for *ev-1*, *ev-2*, and other ASLV show that the regions coding for p19 are highly conserved and that those for p10 are somewhat less conserved. It has been demonstrated previously that the sequences upstream of the *gag* gene, i.e., the leader region, are not highly conserved among several ASLV (2). For example, in the 380 nucleotides of the leader, PrC-RSV differs from *ev-1* by 28, PrC-RSV from SRA-RSV by 39, and SRA-RSV from *ev-1* by 49 single base changes. (In addition, SRA-RSV has an insertion and a deletion in this region). In contrast, *ev-1* and *ev-2* differ at only three positions (Malavarca et al., personal communication), indicating that they are closely related, a conclusion also reached from T1 oligonucleotide mapping of virion RNAs (3). Indeed, all the *ev* loci appear to be closely related to each other by this type of analysis (3). It is therefore surprising that, in one small portion of the part of the *gag* gene that we have sequenced—the C terminus of p19 and p2 and the N terminus of p10—*ev-1* and *ev-2* differ by six nucleotides. These differences, with one exception (nucleotide 974), are clustered between nucleotides 776 and 862, the *ev-2* sequence in this stretch being identical to that of PrC-RSV. One explanation of these observations is that a block of genetic information from an exogenous virus like PrC-RSV was acquired by the progenitor to *ev-2*, presumably before the latter became established in the chicken germ line but only shortly after *ev-1* and *ev-2* diverged from each other. Sequence analysis of the remainder of these genomes and of the other *ev* loci may serve to clarify the phylogenetic relationships of the several chicken endogenous viruses.

ACKNOWLEDGMENTS

This work was supported by Public Health Service grant CA-20081 from the National Cancer Institute.

We thank M. Linial for the quail cell lines and a gift of the radioactive recombinant virus SE21Q1b $\times$ RAV-1, A. Skalka for the DNA clones of *ev-1* and *ev-2* (RAV-0), R. Eisenman for *ev-1*-producing chicken cells, C. Miernicki Steeg for critical reading of the manuscript, and R. Katz for communicating unpublished results. We also acknowledge discussions with S. Oroszlan on the nature of the *gag* protein cleavage sites.

LITERATURE CITED

1. Bhowan, A. S., J. C. Bennett, J. E. Mole, and E. Hunter. 1981. Purification and characterization of the *gag* gene products of avian-type C retroviruses by high pressure liquid chromatography. *Anal. Biochem.* **112**:128–134.
2. Bizub, D., R. A. Katz, and A. M. Skalka. 1984. Nucleotide sequence of noncoding regions in Rous-associated virus-2: comparisons delineate conserved regions important in replication and oncogenesis. *J. Virol.* **49**:557–565.
3. Coffin, J. M., P. N. Tsichlis, K. F. Conklin, A. Senior, and H. L. Robinson. 1983. Genomes of endogenous and exogenous avian

- retroviruses. *Virology* 126:51-72.
4. Conklin, K. F., J. M. Coffin, H. L. Robinson, M. Groudine, and R. Eisenman. 1982. Role of methylation in the induced and spontaneous expression of the avian endogenous virus *ev-1*: DNA structure and gene products. *Mol. Cell. Biol.* 2:638-652.
 5. Cullen, B. R., A. M. Skalka, and G. Ju. 1983. Endogenous avian retroviruses contain deficient promoter and leader sequences. *Proc. Natl. Acad. Sci. USA* 80:2946-2950.
 6. Eisenman, R. N., and V. M. Vogt. 1978. The biosynthesis of oncovirus proteins. *Biochim. Biophys. Acta* 473:187-239.
 7. Fletcher, P., R. C. Nowinski, E. Tress, and E. Fleissner. 1975. Chromatographic separation and antigenic analysis of proteins of the Oncornaviruses. *Virology* 64:358-366.
 8. Gebhardt, A., J. V. Bosch, A. Ziemiecki, and R. R. Friis. 1984. Rous sarcoma virus p19 and gp37 can be chemically cross-linked to high molecular weight complexes. *J. Mol. Biol.* 174:297-317.
 9. Hampe, A., I. Laprevotte, F. Galibert, L. A. Fedele, and C. J. Sherr. 1982. Nucleotide sequences of feline retroviral oncogenes (*v-fes*) provide evidence for a family of tyrosine-specific protein kinase genes. *Cell* 30:775-785.
 10. Herman, A. C., R. W. Green, D. P. Bolognesi, and T. C. Vanaman. 1975. Comparative properties of avian oncornavirus polypeptides. *Virology* 64:339-348.
 11. Hunter, E., J. C. Bennett, A. Bhowan, R. B. Pepinsky, and V. M. Vogt. 1983. Amino-terminal amino acid sequence of p10, the fifth major *gag* polypeptide of avian sarcoma and leukemia viruses. *J. Virol.* 45:885-888.
 12. Johnson, S. P., M. Veigl, T. Vanaman, and J. Leis. 1983. Cyanogen bromide digestion of the avian myeloblastosis virus pp19 protein: isolation of an amino-terminal peptide that binds to viral RNA. *J. Virol.* 45:876-881.
 13. Kitamura, N., A. Kitamura, K. Toyoshima, Y. Hirayama, and M. Yoshida. 1982. Avian sarcoma virus Y73 genome sequence and structural similarity of its transforming gene product to that of Rous sarcoma virus. *Nature (London)* 297:205-208.
 14. Leis, J. P., J. McGinnis, and R. W. Green. 1978. Rous sarcoma virus p19 binds to specific double-stranded regions of viral RNA: effect of p19 on cleavage of viral RNA by RNase III. *Virology* 84:87-98.
 15. Linial, M., E. Medeiros, and W. S. Hayward. 1978. An avian oncovirus mutant (SE21Q1b) deficient in genomic RNA: biological and biochemical characterization. *Cell* 15:1371-1381.
 16. Maxam, A. M., and W. Gilbert. 1977. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* 65:499-560.
 17. O'Rear, J. J., and H. M. Temin. 1982. Spontaneous changes in nucleotide sequence in proviruses of spleen necrosis virus, an avian retrovirus. *Proc. Natl. Acad. Sci. USA* 79:1230-1234.
 18. Palmiter, R. D., J. Gagnon, V. M. Vogt, S. Ripley, and R. N. Eisenman. 1978. The NH₂-terminal sequence of the avian oncovirus *gag* precursor polypeptide (Pr76^{gag}). *Virology* 91:423-433.
 19. Pepinsky, R. B. 1983. Localization of lipid-protein and protein-protein interactions within the murine retrovirus *gag* precursor by a novel peptide mapping technique. *J. Biol. Chem.* 258:11229-11235.
 20. Pepinsky, R. B., and V. M. Vogt. 1979. Identification of retrovirus matrix proteins by lipid-protein cross-linking. *J. Mol. Biol.* 131:819-837.
 21. Pepinsky, R. B., and V. M. Vogt. 1983. Purification and properties of a fifth major viral *gag* protein from avian sarcoma and leukemia viruses. *J. Virol.* 45:648-658.
 22. Pepinsky, R. B., and V. M. Vogt. 1984. Fine-structure analyses of lipid-protein and protein-protein interactions of *gag* protein p19 of the avian sarcoma and leukemia viruses by cyanogen bromide mapping. *J. Virol.* 52:145-153.
 23. Reddy, E. P., R. K. Reynolds, D. K. Watson, R. A. Schultz, J. Lautenberger, and T. S. Papas. 1983. Nucleotide sequence analysis of the proviral genome of avian myelocytomatosis virus (M29). *Proc. Natl. Acad. Sci. USA* 80:2500-2504.
 24. Schwartz, D., R. Tizard, and W. Gilbert. 1983. Nucleotide sequence of Rous sarcoma virus. *Cell* 32:853-869.
 25. Schwartz, W. E., P. K. Smith, and G. P. Royer. 1980. N-(β -iodoethyl) trifluoroacetamide: a new reagent for aminoethylation of thiol groups in proteins. *Anal. Biochem.* 106:43-48.
 26. Seiki, M., S. Hattori, Y. Hirayama, and M. Yoshida. 1983. Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. *Proc. Natl. Acad. Sci. USA* 80:3618-3622.
 27. Shaikh, R., M. Linial, S. Brown, A. Sen, and R. Eisenman. 1979. Recombinant avian oncoviruses. II. Alterations in the *gag* proteins and evidence for intragenic recombination. *Virology* 92:463-481.
 28. Shibuya, M., and H. Hanafusa. 1982. Nucleotide sequence of Fujinami sarcoma virus: evolutionary relationship of its transforming gene with transforming genes of other sarcoma viruses. *Cell* 30:787-795.
 29. Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe. 1981. Nucleotide sequence of Moloney murine leukemia virus. *Nature (London)* 293:543-548.
 30. Swanstrom, R., H. E. Varmus, and J. M. Bishop. 1982. Nucleotide sequence of the 5' noncoding region and part of the *gag* gene of Rous sarcoma virus. *J. Virol.* 41:535-541.
 31. Tung, J. S., and C. A. Knight. 1972. Relative importance of some factors affecting the electrophoretic migration of proteins in sodium dodecyl sulfate-polyacrylamide gels. *Anal. Biochem.* 48:153-163.
 32. Versteegen, R. J., T. D. Copeland, and S. Oroszlan. 1982. Complete amino acid sequence of the group-specific antigen gene-encoded phosphorylated proteins of mouse leukemia viruses. *J. Biol. Chem.* 257:3007-3013.
 33. Vogt, V. M., R. Eisenman, and H. Diggelmann. 1975. Generation of avian myeloblastosis virus structural proteins by proteolytic cleavage of a precursor polypeptide. *J. Mol. Biol.* 96:471-493.
 34. Vogt, V. M., A. Wight, and R. Eisenman. 1979. *In vitro* cleavage of avian retrovirus *gag* proteins by viral protease p15. *Virology* 98:154-167.