JOURNAL OF VIROLOGY, Apr. 1986, p. 50–58
0022-538X/86/040050-09$02.00/0

Vol. 58, No. 1

Structure and Processing of the p2 Region of Avian Sarcoma and Leukemia Virus *gag* Precursor Polyproteins

R. BLAKE PEPINSKY,[1]* ROBERT J. MATTALIANO,[1] AND VOLKER M. VOGT[2]

Department of Protein Chemistry, Biogen Research Corp., Cambridge, Massachusetts 02142,[1] and Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, New York 14853[2]

Received 25 September 1985/Accepted 25 November 1985

We have purified two low-molecular-weight polypeptides from the Prague C strain of Rous sarcoma virus and have identified these as products of the *gag* precursor Pr76 by protein sequencing and by amino acid analysis. Both polypeptides are derived from a stretch of 22 amino acids within Pr76 that separates p19 and p10. We refer to this region as p2. Together the two cleavage products form the entire p2 region. The junctions of p19 with the amino-terminal fragment of p2 and of p10 with the carboxy-terminal fragment of p2 define two new processing sites within the *gag* precursor, Tyr-155–His-156 and Gly-177–Ser-178. Both polypeptides are major cleavage products of Pr76 that occur in Prague C Rous sarcoma virus at an estimated 1,000 copies per virion. They also are prominent components of avian myeloblastosis virus. The combination of gel filtration and reverse-phase high-pressure liquid chromatography, which was used for the isolation of the two fragments of p2, resolved over a dozen other low-molecular-weight polypeptides from avian sarcoma and leukemia viruses that previously were undetected. This technique thus should serve as a useful procedure for further characterization of viral components.

In avian sarcoma and leukemia viruses (ASLV) the *gag* precursor polyprotein Pr76 is processed proteolytically into the five mature proteins (p19, p10, p27, p12, and p15) that form the internal structure of the virus particle. Of these products, p19 is associated with the viral envelope (9, 15), p10 is within the space between the membrane and core (16), p27 forms the core shell (4, 24), and p12 is within the core associated with RNA in a ribonucleoprotein complex (4, 6). The location of p15, the protease responsible for cleaving the viral polyproteins during assembly (7, 26), is unknown. Proteolytic processing of the *gag* precursor also generates minor cleavage products such as p19f (22) and p23 (17) that are less well characterized and that vary in amount in different viruses and virus preparations. Both p19f and p23 are related proteins that differ from p19 only at their carboxy termini. p19f is a 10-kilodalton (kDa) proteolytic fragment derived from the amino-terminal portion of p19, and p23 is an incomplete processing product of Pr76 that contains all of p19 plus p2 (25; see below).

At two positions within the *gag* precursor—the junctions of p19 (carboxy terminus at amino acid 155; 25) with p10 (amino terminus at amino acid 178; 13) and p27 (carboxy terminus at amino acid 479; 1) with p12 (amino terminus at amino acid 489; K. S. Misono, F. S. Sharief, and J. P. Leis, Fed. Proc. **60**:1611, 1980)—the ends of the mature proteins are not contiguous with each other, but are separated by short stretches of amino acids. The region between p19 and p10, which we refer to as p2, is composed of 22 amino acid residues. To determine the fate of this polypeptide in the mature virus, we used a combination of gel filtration and high-pressure liquid chromatography (HPLC) to isolate small polypeptides from virus particles, which then were characterized by amino acid analysis and by protein sequencing. In this report we describe the isolation and characterization of two polypeptides derived from the p2 region, a cysteine-rich fragment from its amino terminus and a proline-rich fragment from its carboxy terminus. Both are prominent species that occur in roughly equimolar amounts in mature virus particles. Based on their presence in both avian myeloblastosis virus (AMV) and the Prague C strain of Rous sarcoma virus (PrC-RSV), we infer that they are major cleavage products of the ASLV *gag* precursor polyprotein generated during virus maturation.

* Corresponding author.

MATERIALS AND METHODS

Virus. Chicken embryo fibroblasts infected either with PrC-RSV or with AMV were grown as monolayers at 39°C in Dulbecco modified Eagle medium containing antibiotics (penicillin, 50 U/ml; streptomycin, 50 μg/ml) and 10% fetal calf serum. Metabolically labeled viruses were purified from the supernatants of infected cells as described previously (17). Unlabeled AMV from leukemic chicken plasma was obtained from Life Sciences Inc. Unlabeled PrC-RSV, in the form of a concentrated suspension from supernatants of virus-infected cell cultures, was obtained from the Biological Carcinogenesis Branch of the National Cancer Institute. Proteins were identified by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) as previously described (16).

Carboxymethylation of viral proteins. For each analysis presented, viral proteins were first carboxymethylated with iodoacetic acid. Briefly, each preparation of virus was suspended at a protein concentration of 5 mg/ml in 200 μl of reducing solution (6 M guanidine hydrochloride, 0.2 M *N*-ethylmorpholine acetate [pH 8.6], 5 mM dithiothreitol) and incubated at 37°C for 2 h. Fresh iodoacetic acid was added to a final concentration of 10 mM (from a 10× stock solution in 0.2 M *N*-ethylmorpholine), and the samples were incubated at 23°C for 30 min in the dark. Unreacted reagent was quenched with 25 mM dithiothreitol. In instances where protein was to be labeled simultaneously, the reduced preparation was first exposed to 250 μCi of iodo[^{3}H]acetic acid (193 mCi/mmol; New England Nuclear Corp.) for 10 min, and then iodoacetic acid was adjusted to a final concentration of 10 mM for the remainder of the reaction period.

HPLC analysis. Viral proteins were subjected to reverse-

phase HPLC at 39°C using the specific loading and elution conditions described for each individual application. Bound components were eluted with 100-min gradients at a flow rate of 1.4 ml/min. For each analysis, 200 0.5-min fractions were collected and either analyzed directly by scintillation counting or stored at −20°C for subsequent analysis. The two HPLC columns that were used in these analyses, C4 (Vydac; catalog no. 214TP54) and C18 (SpectraPhysics, octadecylsilyl; catalog no. A2351-040), are both 0.46 cm (inside diameter) by 25 cm.

Protein sequence analysis. Samples containing either 1 nmol of peptide or 50,000 cpm of peptide in combination with 3 nmol of an unlabeled carrier protein were subjected to amino-terminal sequence analysis by sequential Edman degradation on an Applied Biosystems 470A gas-phase sequencer in the presence of polybrene. Phenylthiohydantoin-amino acids from each cycle were analyzed by reverse-phase HPLC on a 5-μm cyano column (25 by 0.46 cm; IBM Instruments, Wallingford, Conn.) as described (12). The column effluent was monitored spectrophotometrically at both 254 and 313 nm. For radioactive polypeptides, each cycle of the Edman degradation was subjected to scintillation counting.

RESULTS

Reverse-phase HPLC analysis of ASLV proteins. The major proteins in ASLV are the structural proteins that are derived from the *gag* and *env* genes. The five internal structural proteins (*gag* proteins) account for about 80% of the total viral protein, while the two surface glycoproteins (*env* proteins) make up about 20%. When carboxymethylated proteins from PrC-RSV were resolved by reverse-phase HPLC on a C4 column, we obtained the simple absorbance profile shown in Fig. 1A. Proteins from each of the peak fractions were identified by SDS-PAGE. The five major peaks, designated c, d, f, g, and h in Fig. 1A, contain the five *gag* proteins p12, p10, p15, p19, and p27, respectively. Peak e contains gp85. Other prominent peaks—i (p19), j (p19), and k (p27)—are chromatographic variants of peaks g and h, which vary in relative intensity from analysis to analysis. They are indistinguishable from the major peaks by SDS-PAGE. In fact, by gel analysis each of the three peaks for p19 was found to contain both the phosphorylated and unphosphorylated forms of the protein as well as the two related cleavage products, p19f and p23. gp37 was not identified as a component of any of the prominent peaks.

A_{214} values for each of the peaks that were identified in Fig. 1 are summarized in Table 1. Correcting both for the splitting of p27 and p19 into multiple peaks and for the appropriate size adjustments of the absorbance values of the individual peaks, we conclude that p27, p19, p15, and p12 were recovered roughly in equimolar amounts. On the other hand, the single peak for p10 contained only 0.5 mol equivalents of protein. Similar estimates were obtained for p10 from histidine-labeled AMV (described in a later section). Since all five *gag* proteins are equimolar when virus is analyzed by SDS-PAGE (16), a likely explanation for the observed discrepancy is that other unidentified peaks exist for p10.

In addition to the five major *gag* proteins, proteolytic processing of the *gag* precursor also should generate a 22-amino acid polypeptide that within the precursor spans the region between the carboxy terminus of p19 (25) and the amino terminus of p10 (13). We have designated this region provisionally as p2. The predicted amino acid sequence of p2, based on DNA sequencing (19), contains three cysteine residues, which is 20% of the cysteines in Pr76. Thus p2 should be a major component of virus that is labeled metabolically with radioactive cysteine, even though by size it accounts for only 3% of the precursor. We have used this apparent enrichment from cysteine labeling to aid in the identification and purification of p2 from intact virus. Figure 1B shows the distribution of radioactivity when carboxymethylated proteins from [^{35}S]cysteine-labeled PrC-RSV were subjected to reverse-phase HPLC. Fractions were quantitated by scintillation counting. The single large peak, peak c, contains p12, the major cysteine-containing protein in Pr76. The relative sizes of the other peaks also are consistent with their cysteine content. This result is readily apparent from Table 1, where the observed and predicted values for the cysteine content of each of the viral structural proteins are summarized.

In the column profile from cysteine-labeled PrC-RSV, a single new peak (designated a in Fig. 1B) was consistently observed. We have shown by amino acid analysis and by protein sequencing (see below) that this peak contains a fragment of p2. A corresponding peak was observed in the absorbance profile from unlabeled virus (Fig. 2A), but was minor by its absorbance. Using the values summarized in Table 1, we infer that the cysteine-rich fragment of p2 is a major cleavage product of *gag* precursor, since relative to other *gag* proteins it occurs in virus at approximately 0.5 mol equivalents.

Purification of low-molecular-weight polypeptides. We used a combination of gel filtration chromatography and reverse-phase HPLC to isolate low-molecular-weight proteins from PrC-RSV. Before chromatography, viral proteins were labeled and carboxymethylated with iodo[^{3}H]acetic acid. Figure 2A shows the elution profile from gel filtration chromatography in which viral proteins were resolved on an A.5M resin in 6 M guanidine hydrochloride. Portions of fractions were subjected to SDS-PAGE, and the proteins were visualized either directly by Coomassie staining (Fig. 2B) or by fluorography (Fig. 2C). The peak fractions for gp85, p27, p19, p15, and p12 were fractions 40, 44, 50, 57, and 59, respectively. In these analyses p10 was not identified, presumably because it does not contain cysteine (13) and does not stain well with Coomassie blue (16). From fractions 65 through 100, no low-molecular-weight proteins were detected by SDS-PAGE.

By contrast, over a dozen different low-molecular-weight polypeptides were detected when the same gel filtration fractions were analyzed by reverse-phase HPLC. They were characterized initially by their A_{214} and A_{280} and by carboxymethyl-[^{3}H]cysteine content. HPLC profiles from nine serial sets of gel filtration fractions are presented in Fig. 3. Chromatographs at the left show A_{214}, and those at the right show the distribution of radioactivity from the same chromatography fractions. For each analysis the designated gel filtration fractions (from fraction 75 to 95) were combined and loaded directly on a C18 column. Bound components were eluted with a gradient of acetonitrile from 0 to 75% in 25 mM ammonium acetate (pH 6.5).

The molar amounts of each of the polypeptides detected were estimated, using their apparent size based on gel filtration and their relative absorbance. For two components in particular, one containing cysteine but having no A_{280} value (peak a in profiles 86 through 92) and one having an A_{280} value but no cysteines (peak b in profiles 86 through 92), we estimate that they both occur at approximately 1,000 copies per virion. Thus they were assumed to be products of the *gag* gene. These numbers agree with molar estimates

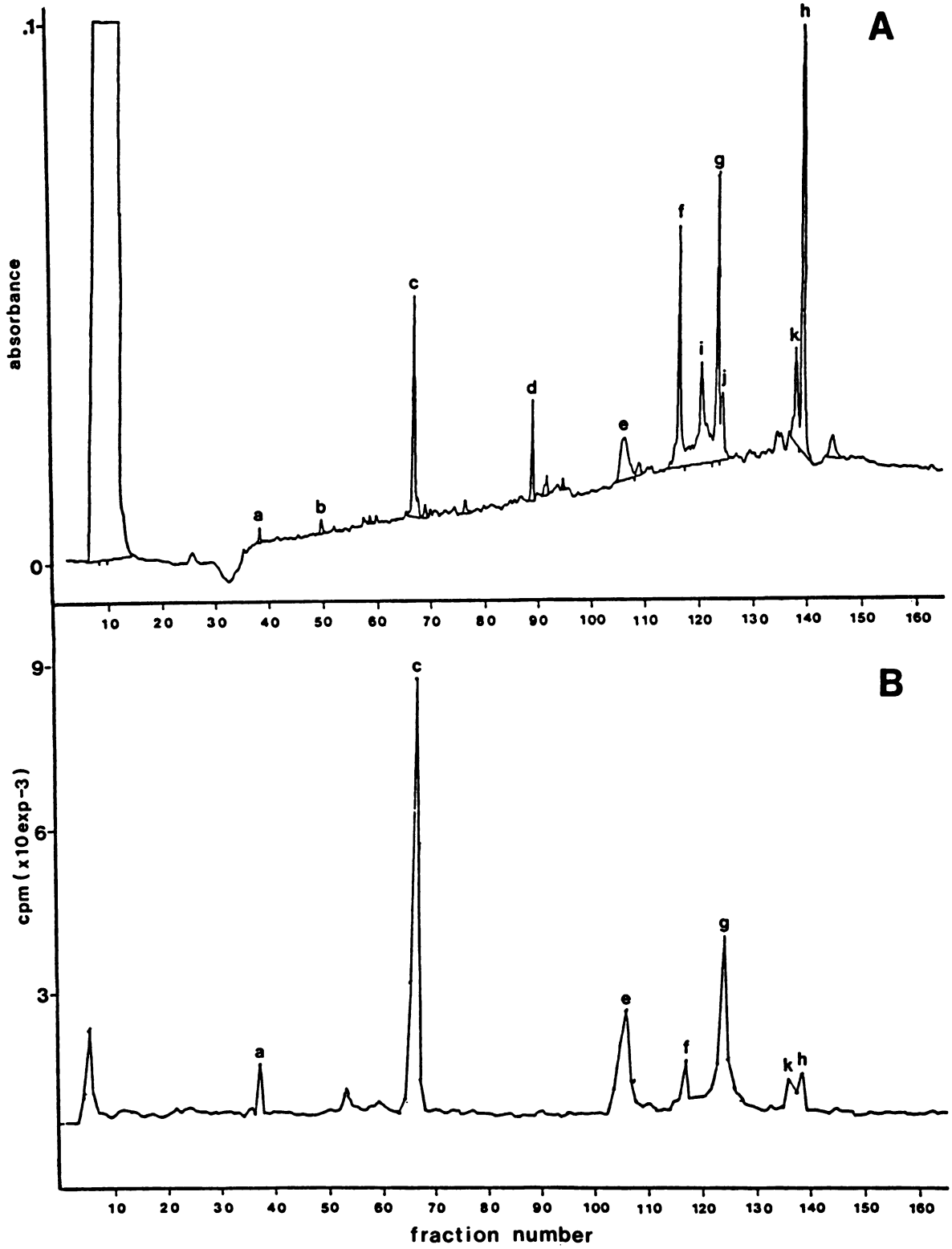

FIG. 1. HPLC analysis of PrC-RSV proteins. Preparations of PrC-RSV, which were carboxymethylated with iodoacetic acid before analysis, were subjected to reverse-phase HPLC on a C4 column. Proteins were eluted with a gradient of acetonitrile from 0 to 75% in 0.1% trifluoroacetic acid. (A) Elution profile from unlabeled virus (50 μg), monitored at 214 nm. (B) Elution profile from [^{35}S]cysteine-labeled virus. In Table 1 the observed data are summarized quantitatively.

TABLE 1. Quantitation of data from Fig. 1

Peak	Protein	No. of amino acids[a]	A_{214}	Mol[b]	No. of cysteines[a]	cpm[c]
gag polypeptides						
a	p2a	9(10)	0.53	0.6(0.5)	2(3)	1,000
b	p2b	11	0.69	0.6	0	0
c	p12	89	9.3	1.0	6	10,800
d	p10	62	3.2	0.5	0	0
f	p15	124	11.3	0.9	1	1,200
g	p19	155	11.5	0.7	4	3,600
h	p27	240	21.0	0.8	1	1,500
env polypeptides						
e	gp85	341	5.9	0.2	14	4,300
—	gp37	198			7	
Other peaks						
i	p19	155	4.2	0.2	4	0
j	p19	155	3.9	0.2	4	540
k	p27	240	5.5	0.2	1	827

[a] Numbers are predicted from the DNA sequence of PrC-RSV (19).
[b] Numbers are calculated from amino acid sequence and A_{214} readings. Values listed are moles relative to p12.
[c] Numbers are derived from the analysis of carboxymethylated, [^{35}S]cysteine-labeled viral proteins. For all values listed, a background of 500 cpm was subtracted.

taken from HPLC profiles of intact virus in Fig. 1. The other low-molecular-weight peptides that were detected occur at much lower copy numbers and remain to be characterized further.

The two prominent low-molecular-weight polypeptides were purified to apparent homogeneity with a second reverse-phase HPLC step. Peak fractions from the initial run were combined, diluted 1:2 with water, and acidified with trifluoroacetic acid. They then were loaded directly onto a C4 column, and bound components were eluted with a gradient of acetonitrile from 0 to 75% in 0.1% trifluoroacetic acid. For the two major peaks described above, we obtained the profiles shown in Fig. 4. Figure 4A is the chromatograph from the major cysteine-containing fragment. All of the tritium label eluted with the single peak at fraction 34. This peptide does not have an A_{280} value. Figure 4B shows the profile derived from the other fragment. It has an A_{280} value but does not have cysteine counts associated with it.

Primary structure of the p2 polypeptides. The cysteine-containing polypeptide shown in Fig. 4A, p2a, was identified as a fragment of p2 by amino acid analysis and by protein sequencing. Before these analyses, two results suggested that this polypeptide was only part of the region. First, the fragment eluted from the gel filtration column as if it were 1 kDa rather than the predicted size of 2.4 kDa based on the sequence of the intact region. Second, the fragment lacked an A_{280} value. The p2 region has a tyrosine near its carboxy terminus. The data in Table 2 summarize the results from amino acid analysis of the cysteine-rich fragment. The observed composition agrees exactly with the predicted composition for a polypeptide derived from the amino-terminal half of p2. It contains His, Cys, Asx, Thr, Gly, Ala, and Ile and is devoid of Lys, Arg, Val, Leu, Phe, Pro, Tyr, and Met. The lower than expected value for cysteine is due to low recovery of carboxymethyl-cysteine by our hydrolysis procedure. In particular the absence of Pro and Tyr, which are unique markers in p2 for the carboxy-terminal half of the region, confirm that only a portion of the molecule is contained in p2a. Low values for Ser and Glu in the analysis presumably result from contaminants.

To further characterize p2a, it was subjected to amino-terminal protein sequence analysis, in which phenylthiohydantoin-amino acids in each cycle of the Edman degredation were analyzed for radioactivity due to tritiated carboxymethyl cysteine. The histogram shown in Fig. 5A summarizes these data. Since radioactivity was released in cycles 2 and 8, we conclude that these positions within the polypeptide contain cysteine. We infer from this result that the fragment starts at His-156 in the predicted protein sequence of Pr76 (19), since such an alignment would place cysteines at positions 2 and 8. This is the only position in the entire sequence of Pr76 where two cysteines are separated by six amino acids. A major implication of this result is that the amino terminus of p2 is immediately adjacent to the carboxy terminus of p19 and thus defines precisely a new cleavage site within the *gag* precursor, Tyr-155–His-156. From the sequence analysis we were unable to identify the carboxy terminus of the peptide; however, based on the amino acid composition presented in Table 2, we can place restrictions on its size. We infer that the fragment must stop at or before the cysteine at position 10 in p2, since it contains Asx (found only at position 9 in p2), but not two Ala (the second Ala in p2 is found at position 11). A polypeptide of this size agrees with the apparent molecular weight of p2a based on gel filtration analysis.

The primary structure of the second prominent low-molecular-weight polypeptide, p2b, was also determined (Fig. 4B). The amino acid composition of this peptide is shown in Table 2. It contains only seven amino acids, Pro, Ala, Ser, Thr, Val, Gly, and Tyr, which suggests that it is derived from the carboxy-terminal half of p2. The fact that the fragment does not contain cysteine implies that its sequence does not overlap with the first peptide. To better evaluate its structure, p2b was subjected to 10 cycles of protein sequencing. From this analysis we obtained the sequence Thr-Ala-Ser-Ala-Pro-Pro-Pro-Pro-Tyr-Val, which agrees exactly with the predicted sequence for the carboxy terminus of p2. Although we did not sequence to the end of the peptide, the fragment presumably ends at Gly-177, based on its amino acid composition. This result implies that the end of p2 is contiguous with p10 and thus defines an additional cleavage site in the *gag* precursor Gly-177–Ser-178.

Together the cysteine- and proline-rich polypeptides make up the entire p2 region. Their organization with respect to the *gag* precursor is shown in Fig. 6. They are major cleavage products of the *gag* precursor that are retained in the mature virus particle. In the preparations of PrC-RSV that were analyzed, there was no evidence of a polypeptide that corresponds to the intact p2 region. However, in our screen for p2, we have not characterized species that occur at less than 10% of the major *gag* proteins. The only other low-molecular-weight cysteine-containing fragment that was subjected to sequence analysis (fraction 54 in profiles 74 through 79 of Fig. 3) had a blocked amino terminus and thus presumably is derived from some other protein (not shown).

p2 polypeptides in AMV. To determine whether the processing pattern for p2 in PrC-RSV is a common phenomena of ASLV or is specific to that virus strain, we also analyzed by HPLC the protein component of AMV. As purified from leukemic chicken plasma, AMV is actually a mixture of myeloblastosis-associated viruses and AMV itself. These closely related viruses have been sequenced only partially, but amino acid analyses (8, 11), immunological properties

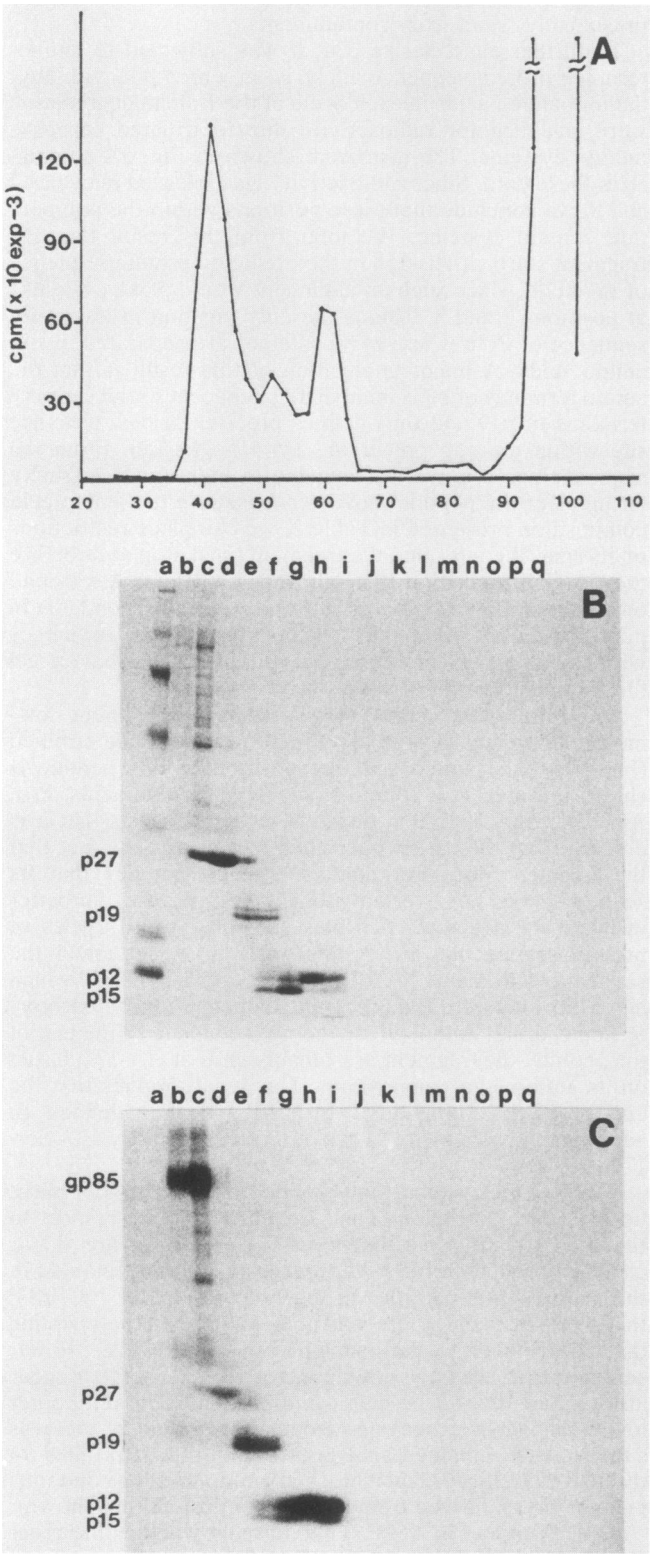

FIG. 2. Gel filtration analysis of PrC-RSV proteins. A 2-mg sample of virus was carboxymethylated with iodo[^{3}H]acetic acid and then subjected to gel filtration on an A.5M column (1 by 40 cm) in 6 M guanidine hydrochloride–0.05 M sodium phosphate (pH 5.4)–10 mM dithiothreitol. Chromatography was preformed at room temperature with a flow rate of 1 ml/h. Fractions of 0.3 ml were collected. Samples of fractions were analyzed either directly by scintillation counting (A) or, after SDS-PAGE, by Coomassie staining (B) or fluorography (C). For gel analysis, proteins were first selectively precipitated from the guanidine hydrochloride with ethanol (16). Lanes b through q correspond to gel filtration fractions 39, 43, 47, 50, 53, 56, 59, 62, 65, 68, 71, 74, 77, 83, 89, and 97, respectively.

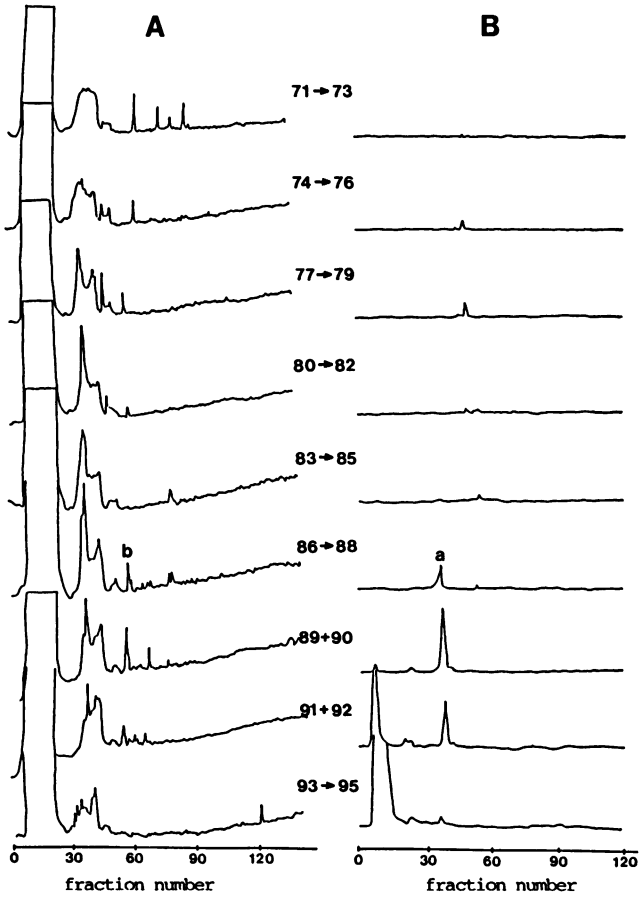

FIG. 3. HPLC chromatographs of low-molecular-weight proteins. Sequential sets of gel filtration samples, which contained low-molecular-weight components from PrC-RSV, were further fractionated by reverse-phase HPLC. The gel filtration fractions indicated in the middle of the figure were combined and loaded directly onto a C18 column. Bound components were eluted with a gradient of acetonitrile from 0 to 75% in 25 mM ammonium acetate (pH 6.5). Profiles at the left (column A) show A_{214}. Profiles at the right (column B) show the distribution of radioactive polypeptides from the same sets of fractions.

(3), and existing sequence data for some of the *gag* proteins (18) suggest that the proteins from PrC-RSV and AMV and its helper viruses are all highly conserved. By SDS-PAGE the apparent sizes of the structural proteins from the two viruses are virtually identical, except for p10, which migrates anomalously in SDS gels (16).

When [^{35}S]cysteine-labeled AMV (carboxymethylated with iodoacetic acid) was subjected to HPLC, we obtained the profile shown in Fig. 7A. The major proteins p12, p10, gp85, p15, p19, and p27 eluted at exactly the same positions in the gradient as the corresponding peaks from PrC-RSV. This result is based on a simultaneous analysis of radioactive

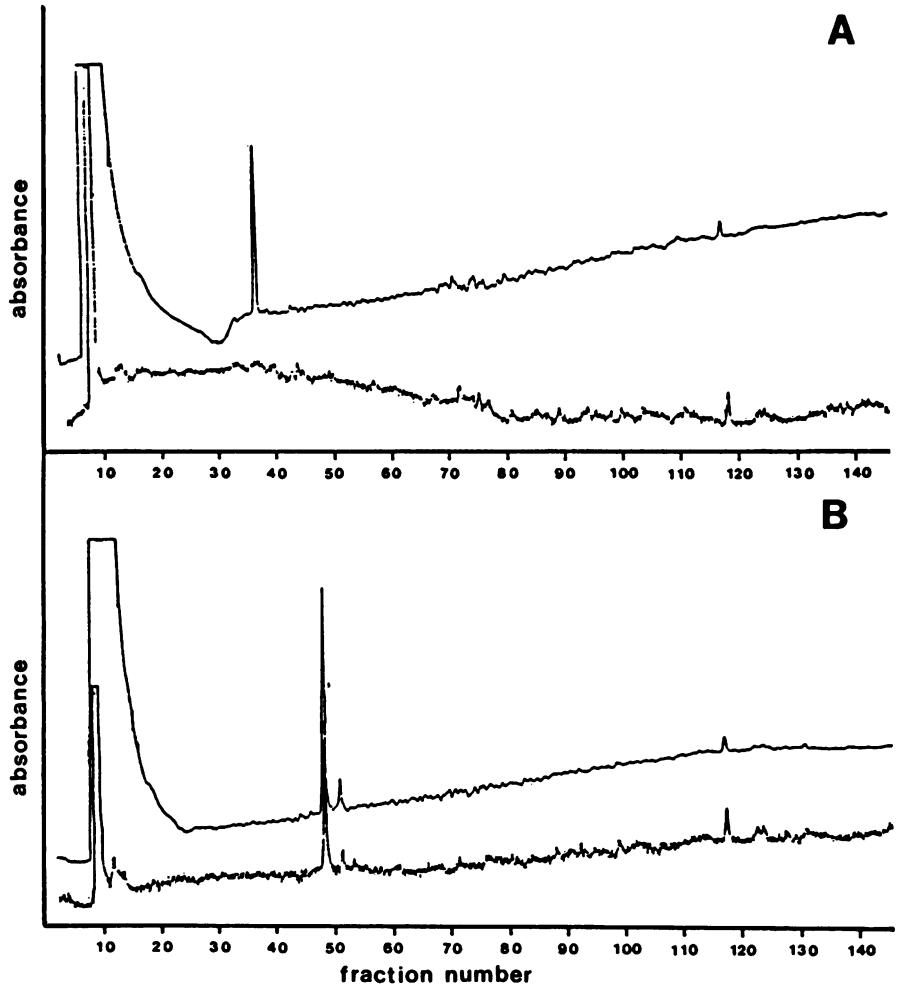

FIG. 4. Reruns of p2 fragments on HPLC. Fractions containing the major low-molecular-weight peptides detected in profiles shown in Fig. 3 were rechromatographed on a C4 column. Bound components were eluted with a gradient of acetonitrile (0 to 75%) in 0.1% trifluoroacetic acid. Fractions were monitored both at 214 nm (upper trace, 0.1 full-scale deflection) and at 280 nm (lower trace, 0.01 full-scale deflection). (A) Column profile of cysteine-rich p2 fragment. (B) Column profile of proline-rich p2 fragment.

AMV with unlabeled PrC-RSV. The single major difference between the profiles was in the region of the cysteine-rich p2 fragment. Whereas a single [^{35}S]cysteine peak was observed in the chromatograph from PrC-RSV, three peaks were observed in the profile from AMV. The first peak coeluted with the cysteine-containing p2 fragment from PrC-RSV, while the two other peaks eluted at slightly higher acetonitrile concentrations. We obtained the same result with AMV that was isolated from chicken cells grown in culture and from leukemic chickens.

To decide whether the additional peaks specific to AMV were also derived from the p2 region, we used HPLC to analyze radioactive AMV labeled metabolically with histidine. The single histidine in p2 accounts for 10% of all the histidines in Pr76, and thus, as with cysteine, metabolic labelings with histidine should generate an apparent enrichment for p2 in virus. When histidine-labeled AMV was subjected to HPLC we obtained the elution profile shown in Fig. 7B. The major peak in the chromatograph corresponds to p15, which is the major histidine-containing *gag* protein in ASLV. Likewise, the sizes of peaks corresponding to each of the other proteins were consistent with their histidine content. p19, which does not contain histidine (25), was not detected in the chromatograph. As observed with the cysteine-labeled AMV, there were multiple peaks in the region of the p2 fragments, suggesting that the peptides in these additional peaks are derived from the p2 region.

The two largest peaks in the triplet were analyzed by protein sequencing. In this analysis, proteins from AMV that were carboxymethylated with iodo[^{3}H]acetic acid were resolved by HPLC, and the appropriate peaks were subjected to sequence analysis. Each cycle was analyzed by scintillation counting. The histogram shown in Fig. 5B is derived from sequence analysis of the first of the peaks in the chromatogram. Radioactivity was detected in cycles 2, 8, and 10. Thus the peak is derived from p2. However, unlike preparations from PrC-RSV, the protein sequence of the AMV protein has a phasing problem that is apparent in later sequencing cycles. As a result of this problem, cycle 9 has more counts than cycle 8 and cycle 11 is larger than cycle 10. Such a phasing problem is often caused by a proline in the peptide sequence and suggests that the AMV peptide may differ slightly in sequence from the corresponding protein in PrC-RSV. However, the sequence difference was not significant enough to change the peptide's elution pattern from HPLC.

For the second AMV peptide sequenced we obtained the same histogram that is shown in Fig. 5B. Thus, both peaks

contain the amino terminus of the p2 region. Although we were unable to determine the source of the heterogeneity, neither peak appears to be intact p2 since they both lacked A_{280}. Furthermore, a peak was detected in the absorbance profile that corresponds to the C-terminal fragment of p2 in both its elution position and the fact that it has an A_{280} value. Based on our results from PrC-RSV and AMV, we conclude that the proteolytic processing that we have characterized for p2 is a common pathway in the processing of the ASLV *gag* precursor polyprotein.

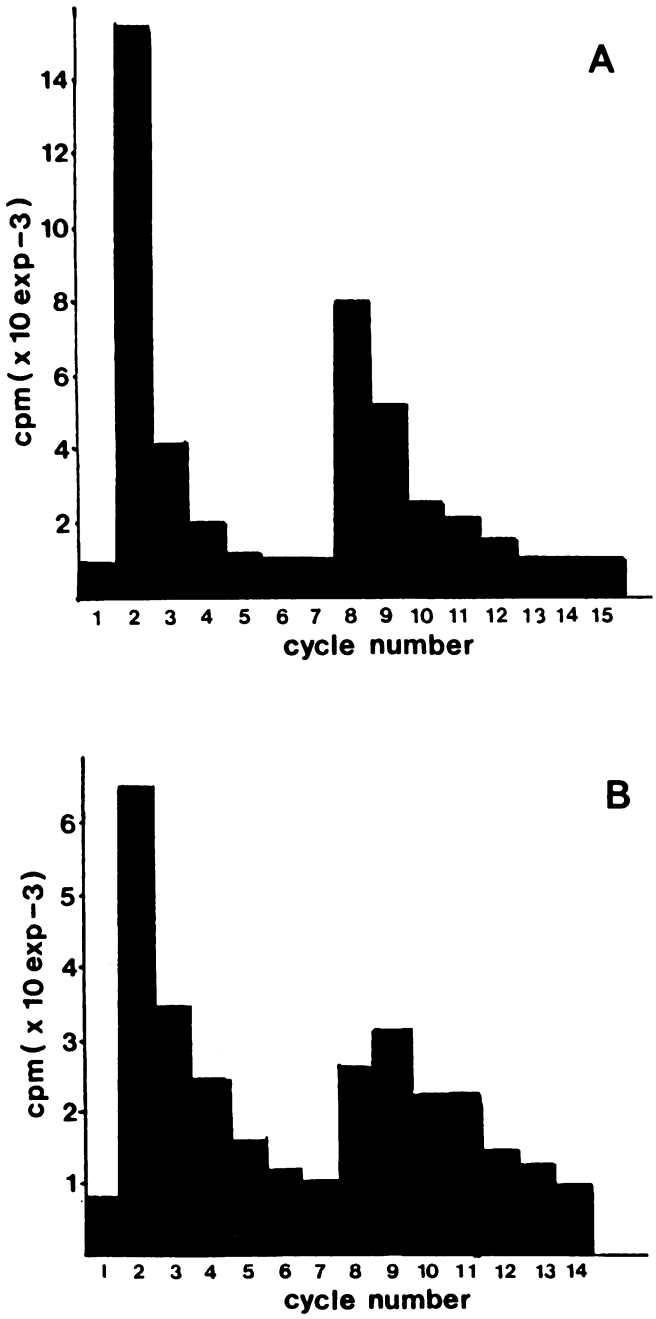

FIG. 5. Sequence analysis of cysteine-labeled p2 from PrC-RSV (A) and AMV (B). Preparations of HPLC-purified p2, which were carboxymethylated with iodo[^{3}H]acetic acid, were subjected to protein sequencing. Each cycle was analyzed for radioactivity by scintillation counting. The histograms shown summarize the results from these analyses.

TABLE 2. Amino acid analysis of p2 polypeptides[a]

Amino acid	Composition of p2 sequences from:			
	p2a		p2b	
	PrC-RSV DNA	PrC protein	PrC-RSV DNA	PrC protein
Aspartic acid	1	1.5	0	0.1
Threonine	1	1.2	1	0.9
Serine	0	0.4	1	1.1
Glutamic acid	0	0.3	0	0.2
Proline	0	0	4	3.9
Glycine	2	2.2	1	1.2
Alanine	1	1	2	2
Cysteine	2(3)	1.4	0	0
Valine	0	0	1	1.3
Methionine	0	0	0	0
Isoleucine	1	1	0	0
Leucine	0	0.2	0	0
Tyrosine	0	0	1	0.8
Phenylalanine	0	0	0	0
Histidine	1	0.8	0	0
Lysine	0	0	0	0
Arginine	0	0	0	0

[a] For amino acid compositions, 250 pmol of each peptide was hydrolized in 6 N HCl (Pierce Chemicals) at 110°C for 24 h in sealed evacuated ampoules and then analyzed for amino acid content on a Beckman System 6300 analyzer. Values listed are moles of each amino acid per mole of fragment. Zeros denote values that were less than 0.05. Predicted values based on DNA sequencing are for the 22-amino acid sequence that extends from the C-terminus of p19 to the amino terminus of p10.

DISCUSSION

We have purified two low-molecular-weight polypeptides from PrC-RSV that were identified as fragments of the *gag* precursor polyprotein by partial protein sequencing. They are derived from a stretch of 22 amino acids within Pr76 that separates p19 and p10. We refer to this region of the precursor as p2. Together the two polypeptides form the entire p2 region. The sequences of the two fragments and their relationship to the *gag* precursor are summarized in Fig. 6. p2a is a cysteine-rich nonapeptide from the amino terminus of p2. It is contiguous with the carboxy terminus of p19. p2b is a proline-rich dodecapeptide from the carboxy terminus of p2 and is contiguous with the amino terminus of p10. From our analyses there was no evidence of an intact polypeptide in virus that corresponds to the entire p2 region;

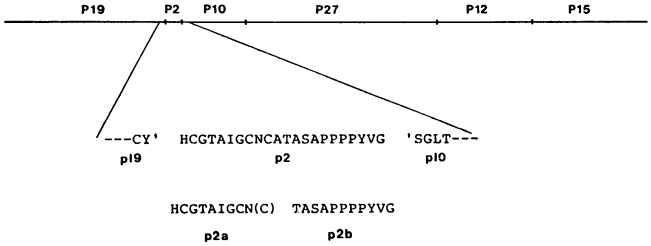

FIG. 6. Summary of structural analyses of PrC-RSV polypeptides. The top line is a schematic representation of the ASLV *gag* precursor polyprotein Pr76. In the diagram, each of the mature structural proteins is indicated with respect to its position within the precursor. Below the line, the p2 region has been expanded. The predicted amino acid sequence for this region, based on the DNA sequence of PrB-RSV, is shown. Below this sequence are the protein sequences for the two *gag* polypeptides, p2a and p2b.

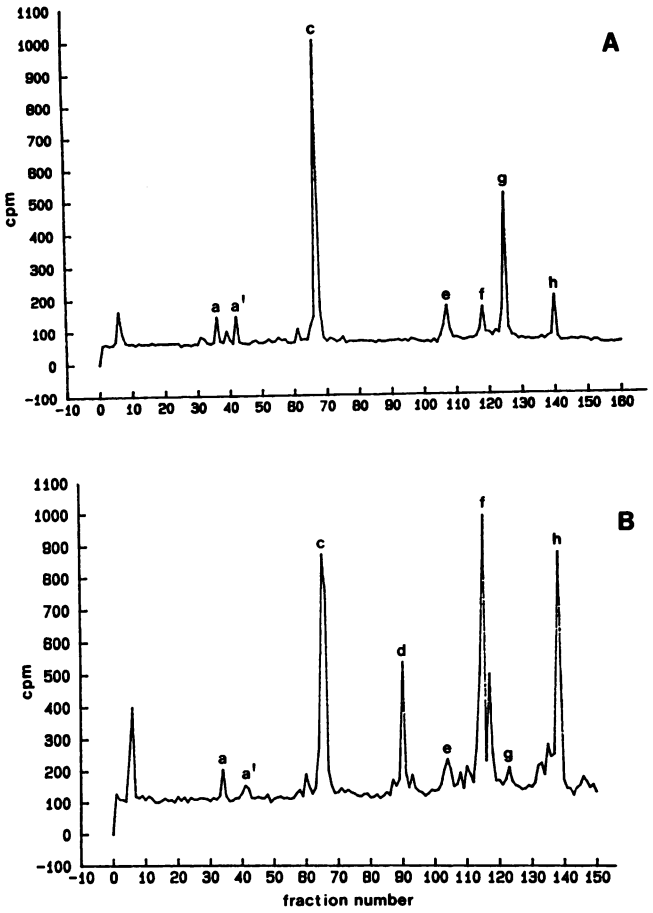

FIG. 7. HPLC analysis of AMV proteins. Preparations of carboxymethylated proteins from AMV were subjected to reverse-phase HPLC on a C4 column as described in the legend of Fig. 1. Each fraction was analyzed by scintillation counting. (A) Elution profile from [^{35}S]cysteine-labeled AMV. (B) Elution profile from [^{3}H]histidine-labeled AMV. The letter designations for each peak are the same designations as in Fig. 1. a' is the single peak that is specific to AMV.

however, we have not characterized any species that occur at less than 10% of the major *gag* proteins.

The two p2 fragments are major cleavage products of Pr76, which occur at an estimated 1,000 copies per virion. Their identification signifies a set of additional processing events that previously was not recognized. Furthermore, the junctions of p2 with existing mature *gag* proteins define two new cleavage sites within Pr76, the p19-p2 junction (Tyr-His) and the p2-p10 junction (Gly-Ser). The same fragments also were detected in preparations of AMV, suggesting that they represent a common cleavage pathway in the maturation of ASLV *gag* proteins. However, we were unsuccessful in attempts to generate the peptides by in vitro processing of Pr76 with the purified avian protease p15 (not shown).

In addition to the two major low-molecular-weight fragments, some of the p2 amino acid sequences in virus are retained on longer polypeptides due to incomplete processing of the *gag* precursor. The most prominent of these products is referred to as p23, which consists of p19 plus an extra stretch of amino acid residues at its carboxy terminus (17). It is likely that p23 contains the entire p2 region (25). In PrC-RSV, p23 accounts for about 10% of the p19 species, but in other ASLV, such as recombinants between RSV and endogenous viruses, p23 can account for as much as 50% of the p19 species (21). In addition to p23, in several virus preparations we have detected cysteine associated with a p10-related polypeptide. This species also must be derived from p2, since the first cysteine in p27 is 20 kDa into the protein (unpublished data).

The data for p2 show unequivocably that the previous assignment of the carboxy terminus of p19 to Tyr-175 (2) is incorrect. This tyrosine is contained within the proline-rich fragment of p2 that we sequenced. We had inferred this result previously from studies in which p19 was analyzed by carboxypeptidase digestion and by metabolic labeling (25). Similarly, our present analyses also show that His-156 is in p2 and not in p19 as previously suggested (2). It is not clear why all existing amino acid analyses for p19 contain at least a part of a histidine (2, 8, 11). Perhaps the false readings for histidine in these analyses result from a modified amino acid residue within p19 that elutes at the position of histidine in amino acid analyses.

The p2 region is highly conserved in ASLV. In seven different viruses that were evaluated by DNA sequencing (summarized in reference 25), there were only two mutations in this region that resulted in amino acid changes. Part of the p2 amino acid sequence is conserved in other retroviruses apparently unrelated to ASLV. The sequence Pro-Pro-Pro-Tyr is found in approximately the same location in the *gag* precursor of mammalian retroviruses such as murine leukemia virus (23), feline leukemia virus (10), and human T-cell leukemia virus type I (20) that bear no other sequence homology to ASLV. It is also found in the avian spleen necrosis virus (14). However, in none of these other viruses is there any evidence that small polypeptides are generated from this portion of the *gag* gene. Recently it was shown that the p2 region of the *gag* gene may be translated in a second reading frame from a differently spliced mRNA as part of a 12-kDa protein (5). This protein was suggested to function as a transcriptional regulator. The existence of two functional proteins translated in different frames from the same portion of the genome would impose severe evolutionary constraints on nucleic acid divergence. The unusual conservation of the Pro-Pro-Pro-Tyr sequence might result from such constraints. It is also possible that this amino acid sequence is conserved because of a *cis*-acting signal in the nucleic acid sequence that encodes it, such as an enhancer sequence.

The approach that we used for isolation of the p2 fragments, i.e., the combination of gel filtration and reverse-phase HPLC, should prove to be a useful method for the isolation of other low-molecular-weight components from virus. In the simple analysis presented, we detected over a dozen polypeptides from which we analyzed only the two prominent *gag* cleavage products. Analysis of the other components should provide further insight into the processing events involved in virus maturation.

ACKNOWLEDGMENTS

We thank E. Pingchang Chow for expert technical assistance for both analysis of amino acid compositions and HPLC analysis.

V.M.V. is supported by Public Health Service grant CA-20081 from the National Cancer Institute.

LITERATURE CITED

1. **Bhown, A. S., J. C. Bennett, and E. Hunter.** 1980. Alignment of the peptides derived from acid-catalyzed cleavage of an aspartylprolyl bond in the major internal structural polypeptide of avian retroviruses. J. Biol. Chem. **255**:6962–6965.

2. **Bhown, 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.
3. **Bolognesi, D. P., R. Ishizaki, G. Huper, T. C. Vanaman, and R. E. Smith.** 1975. Immunological properties of avian oncornavirus polypeptides. Virology **64**:349–357.
4. **Bolognesi, D. P., R. Luftig, and J. H. Sharper.** 1973. Localization of RNA tumor virus polypeptides. I. Isolation of further virus substructures. Virology **56**:549–564.
5. **Broome, S., and W. Gilbert.** 1985. Rous sarcoma virus encodes a transcriptional activator. Cell **40**:537–546.
6. **Davis, N. L., and R. R. Rueckert.** 1972. Properties of a ribonucleoprotein particle isolated from Nonidet P40-treated Rous sarcoma virus. J. Virol. **10**:1010–1020.
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. **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.
9. **Gebhardt, A., J. V. Bosch, A. Ziemiecki, and R. R. Frijs.** 1984. Rous sarcoma virus p19 and gp85 can be chemically cross-linked to high molecular weight complexes. J. Mol. Biol. **174**:297–317.
10. **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.
11. **Herman, A. C., R. W. Green, D. P. Bolognesi, and T. C. Vanaman.** 1975. Comparative properties of avian oncornavirus polypeptides. Virology **64**:339–348.
12. **Hunkapillar, M., and L. Hood.** 1983. Analysis of phenylthiohydantoins by ultrasensitive gradient high performance liquid chromatography. Methods Enzymol. **91**:486–493.
13. **Hunter, E., J. C. Bennett, A. Bhown, 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.
14. **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.
15. **Pepinsky, R. B., and V. M. Vogt.** 1979. Identification of retrovirus matrix proteins by lipid-protein cross-linking. J. Mol. Biol. **131**:819–837.
16. **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.
17. **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.
18. **Sauer, R. T., D. W. Allen, and H. D. Niall.** 1981. Amino acid sequence of p15 from avian myeloblastosis virus complex. Biochemistry **20**:3784–3791.
19. **Schwartz, D., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32**:853–869.
20. **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.
21. **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.
22. **Shealy, D. J., A. G. Mosser, and R. R. Ruekert.** 1980. Novel p19-related protein in Rous-associated virus type 61: implications for avian *gag* gene order. J. Virol. **34**:431–437.
23. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukemia virus. Nature (London) **293**:543–548.
24. **Stromberg, K., N. E. Hurley, N. L. Davis, R. R. Rueckert, and E. Fleissner.** 1974. Structural studies of avian myeloblastosis virus: comparison of polypeptides in virion and core component by dodecyl sulfate-polyacrylamide gel electrophoresis. J. Virol. **13**:513–528.
25. **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.
26. **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.