JOURNAL OF VIROLOGY, May 1981, p. 581-592
0022-538X/81/050581-12$02.00/0

Vol. 38, No. 2

Structure of Glycosylated and Unglycosylated *gag* Polyproteins of Rauscher Murine Leukemia Virus: Carbohydrate Attachment Sites

ALAN M. SCHULTZ,* SALLY M. LOCKHART, EVELYN M. RABIN, AND STEPHEN OROSZLAN

Biological Carcinogenesis Program, Frederick Cancer Research Center, Frederick, Maryland 21701

Received 2 December 1980/Accepted 13 February 1981

The structural relationships among the *gag* polyproteins Pr65^{gag}, Pr75^{gag}, and gPr80^{gag} of Rauscher murine leukemia virus were studied by endoglycosidase H digestion and formic acid cleavage. Fragments were identified by precipitation with specific antisera to constituent virion structural proteins followed by one-dimensional mapping. Endoglycosidase H reduced the size of gPr80^{gag} to that of Pr75^{gag}. By comparing fragments of gPr80^{gag} and the apoprotein Pr75^{gag}, the former was shown to contain two mannose-rich oligosaccharide units. By comparing fragments of Pr65^{gag} and Pr75^{gag}, the latter was shown to differ from Pr65^{gag} at the amino terminus by the presence of a leader peptide approximately 7,000 daltons in size. The internal and carboxyl-terminal peptides of the two unglycosylated polyproteins were not detectably different. The location of the two N-linked carbohydrate chains in gPr80^{gag} has been specified. One occurs in the carboxyl-terminal half of the polyprotein at asparagine$_{177}$ of the p30 sequence and the other is found in a 23,000-dalton fragment located in the amino-terminal region of gPr80^{gag} and containing the additional amino acid sequences not found in Pr65^{gag} plus a substantial portion of p15.

Synthesis of structural proteins as precursor polyproteins is a biosynthetic process common to many animal viruses (5). In murine leukemia viruses (MuLV), the four internal structural proteins of the virion (p10, p12, p15, and p30) occur together on a polyprotein encoded by the *gag* gene. The proximal precursor to these proteins is a 65,000-dalton (65K) molecule, designated Pr65^{gag} (2, 10, 34).

An unexpected feature of the *gag* gene products in MuLV is that glycosylated forms of the *gag* polyproteins are found (8, 11, 29), although none of the four structural proteins comprising the core of the virus is glycosylated. These glycoproteins have not been found to be associated with viral particles but occur on the cell surface. The biosynthetic and structural relationships among the proteins encoded by the *gag* gene are of interest, especially since the Gross cell surface antigen, a marker antigen in Gross MuLV-induced leukemias (22), has been shown to be a glycosylated *gag* polyprotein, designated gPr95^{gag} (18, 31).

In the case of Rauscher MuLV (R-MuLV), the original investigations into *gag* polyproteins demonstrated the existence of a polyprotein larger than Pr65^{gag} but with a shorter half-life in pulse-chase experiments (2, 34). Subsequently, it was shown that the larger molecule, designated gPr80^{gag}, contained mannose-rich oligosaccharides and served as a precursor to the highly glycosylated gPr94^{gag}, a molecule in R-MuLV analogous to the Gross cell surface antigen (29). Similar results were obtained for Moloney MuLV (Mo-MuLV) (8). It was also shown that, in addition to its carbohydrate content, gPr80^{gag} differed from Pr65^{gag} by containing a long leader peptide, whereas the carboxyl termini of the two polyproteins appeared to be the same and may be identical (28). We have further investigated the structure of gPr80^{gag} to specify the location and number of carbohydrate chains and to ascertain whether glycosylation in gPr80^{gag} occurs on the leader peptide or within the known viral structural protein sequences with gPr80^{gag}.

MATERIALS AND METHODS

Cells and cell culture. JLS-V9 cells (36) chronically infected with R-MuLV were obtained from the Viral Resources Laboratory (Frederick Cancer Research Center, Frederick, Md.). Fetal calf serum was purchased from GIBCO Laboratories (Grand Island, N.Y.). Eagle minimal essential medium (EMEM) and methionine-free EMEM were obtained from Flow Laboratories (Rockville, Md.).

Cells were grown on Corning plastic ware in EMEM supplemented with 10% heat-inactivated fetal calf serum, penicillin (100 IU/ml), and streptomycin (100 μg/

ml) under an atmosphere of 5% CO_2 in air. Cultures were regularly tested for mycoplasma contamination by R. Del Giudice (Frederick Cancer Research Center) and were maintained mycoplasma-free.

Radioisotopes and reagents. [^{35}S]methionine (400 to 900 Ci/mmol) and 2,5-diphenyloxazole were purchased from Amersham Corp. (Arlington Heights, Ill.). [^{14}C]leucine (>270 mCi/mmol), [^{14}C]lysine (>270 mCi/mmol), and ^{14}C-labeled protein molecular weight markers were purchased from New England Nuclear Corp. (Boston, Mass.). Protein A-Sepharose was purchased from Sigma Chemical Co. (St. Louis, Mo.). Tunicamycin was obtained from the Lilly Research Laboratory (Indianapolis, Ind.) through the courtesy of Robert Hamill. Endoglycosidases D and H were obtained from Miles Laboratories (Elkhart, Ind.) and tolylsulfonyl phenylalanyl chloromethyl ketone-trypsin was from Worthington Diagnostics (Freehold, N.J.).

Antisera. Antisera to purified R-MuLV p30 and p10 were prepared in goats as previously described (20, 24). Antiserum to R-MuLV p15 prepared in goats was obtained from Roger Wilsnack, Huntingdon Laboratories (Baltimore, Md.). This particular antiserum (lot 4S-716) also has some anti-p12 activity, as indicated in the analytical information accompanying the serum.

Radiolabeling of cell cultures. Cells were labeled either in methionine-free EMEM supplemented with 5% dialyzed fetal calf serum as previously described (29) or in EMEM lacking leucine and lysine. For analytical experiments, [^{35}S]methionine (5 μCi/ml) was incorporated for 15 min after a 10-min preincubation in methionine-free medium. To prepare Pr65^{gag} and gPr80^{gag} for cleavage experiments, labelings were performed as above with [^{35}S]methionine at 40 μCi/ml or for 30 min with leucine and lysine (125 μCi/ml each) after a 10-min preincubation in EMEM lacking leucine and lysine. To label Pr75^{gag} for cleavage experiments, [^{35}S]methionine at 40 μCi/ml was incorporated after a 1-h preincubation with tunicamycin (5 μg/ml), or, similarly, leucine and lysine (125 μCi/ml each) were incorporated after a 1-h preincubation with tunicamycin. To obtain p30 for analysis, a T150 flask of JLS-V9 cells was labeled with [^{35}S]methionine (50 μCi/ml) for 1 h in 5 ml of methionine-free EMEM. Next, 25 ml of complete EMEM was added, and the culture was incubated overnight. Medium was clarified at low speed, and virus particles were recovered by centrifugation for 90 min at 105,000 $\times$ *g* through a 20% sucrose cushion. The virus pellet was subsequently dissolved and electrophoresed as described below for immune precipitates.

Immunoprecipitation and preparation of polyproteins. At the end of a labeling period, cell monolayers were rinsed, lysed, clarified, and immunoprecipitated with anti-p30 serum, using protein A-Sepharose as previously described (28). Washed immunoprecipitates were subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) in 7.5 to 17% linear gradient slabs according to the formulation of Laemmli (15). For subsequent recovery of proteins, bands were located by direct radioautography; otherwise scintillation radioautography was used (16). Radioactive proteins were recovered from dried gel slices by rehydrating and soaking in 0.1% SDS followed by lyophilization in siliconized tubes.

Cleavage of polyproteins. Enzymatic and chemical cleavages of proteins were performed as follows. Endoglycosidase D and H digestions were performed on washed immunoprecipitates as previously described (30). Tryptic digestions were performed on dried gel slices as described for preparations for tryptic maps (28), except that myoglobin and performic acid oxidation were omitted. Formic acid cleavage of polyproteins was performed by dissolving lyophilized gel eluates in 1 ml of 70% formic acid. After standing for 24 h at 25°C, formic acid was removed by repeated lyophilization.

Immunoprecipitation of acid cleavage products. Formic acid-cleaved polyproteins purified as described above were redissolved in 1 ml of lysing buffer (35). The cleavage mixture was first treated with anti-p10 serum, and the immune complexes were removed with protein A-Sepharose (28). The supernatant was then treated with anti-p30 serum and protein A-Sepharose. Similarly, the supernatant from this treatment received anti-p15 serum and protein A-Sepharose. Material recovered at each step in the sequential antibody treatment as a washed protein A-Sepharose precipitate was frozen and stored until electrophoresis.

Gel electrophoresis of tryptic digests. Solubilized peptides released from gel slices by tryptic digestion were separated from the gel and relyophilized several times from water in siliconized tubes. The digests were subjected to PAGE under nondenaturing conditions in 15% slab gels formulated according to Davis (7). [^{35}S]methionine-labeled peptides localized by radioautography of dried gels were eluted from gel slices with 0.1% SDS as described above.

Amino acid sequencing of peptides. Semiautomated analyses utilizing Edman degradation were performed with a Beckman sequencer model 890C as described elsewhere (6). Briefly, purified [^{35}S]methionine-labeled peptides were analyzed in the presence of Polybrene (12) and myoglobin, as carrier, using a dilute Quadrol program as previously described (23). To determine the position of methionine within the sequence, fractions collected at each cycle of the sequencer run were counted in a Beckman LS-230 liquid scintillation counter.

RESULTS

Digestion of *gag* polyproteins with endoglycosidases. The glycosylated *gag* polyprotein gPr80^{gag} is expected to be larger than Pr65^{gag} if only due to its carbohydrate content. Previously, it was shown that gPr80^{gag} also contains peptide sequences not found in Pr65^{gag} because the apoprotein of gPr80^{gag}, synthesized in the presence of tunicamycin, has a molecular weight of 75K (Pr75^{gag}) (29). An independent method of generating an apoprotein free of asparagine-linked carbohydrate is treatment of the glycoprotein with an endoglycosidase. In Fig. 1, it is shown that digestion of gPr80^{gag} with endoglycosidase H also produces Pr75^{gag}; endoglycosidase D is not effective. The susceptibility to

endoglycosidase H is consistent with the presence of immature, mannose-rich oligosaccharide cores in gPr80^{gag} (33). The essential residue for the action of endoglycosidase D is an unsubstituted α-mannosyl residue linked to the innermost β-mannosyl residue by a 1 $\rightarrow$ 3 linkage (32). Endoglycosidase H will cleave whether that mannose is substituted or not, so the structure of the oligosaccharides of gPr80^{gag} is of the more general type usually found (33). The molecular weight differences between gPr80^{gag} and its unglycosylated form, whether generated biosynthetically in the presence of tunicamycin or by enzymatic degradation, suggest that two mannose-rich oligosaccharide units with molecular weights of 2.5K each (19) are present.

Acid cleavage of *gag* polyproteins: amino-terminal fragments. Treatment of *gag* polyproteins Pr65^{gag} and gPr80^{gag} with formic acid causes chain scission at an acid-sensitive site within the p12 sequence to yield two large fragments as major products. As shown previously, the molecular weight of the parent molecule is accounted for by adding the molecular weights of the two major cleavage fragments, one extending from the amino terminus of the polyprotein to the acid-sensitive site in the p12 and the other extending from the p12 cleavage site to the carboxyl terminus of the polyprotein (28). This technique has been extended in the present study to analyze Pr75^{gag} which has been synthesized in the presence of tunicamycin. Size comparison of fragments obtained from gPr80^{gag} and Pr75^{gag} should determine whether one or both of the two large fragments contain carbohydrate. Acid cleavage fragments are identified by reactivity with specific antisera and by size as previously described (28) (Fig. 2). The fragments, identified by their assigned letters, are summarized in Table 1, along with molecular weight estimates from Fig. 2. Because the gene order of the individual viral proteins in Pr65^{gag} is known to be NH_2-p15-p12-p30-p10-COOH (4, 27), cleavage fragments can be aligned in the Pr65^{gag} sequence. Cleavage products of gPr80^{gag} occur in lanes 1 through 3, those of Pr65^{gag} occur in lanes 4 through 6, and those of Pr75^{gag} occur in lanes 7 through 9. Immune precipitation was done first with anti-p10 serum (lanes 1, 4, and 7), followed by anti-p30 serum (lanes 2, 5, and 8), and finally by anti-p15 serum (lanes 3, 6, and 9).

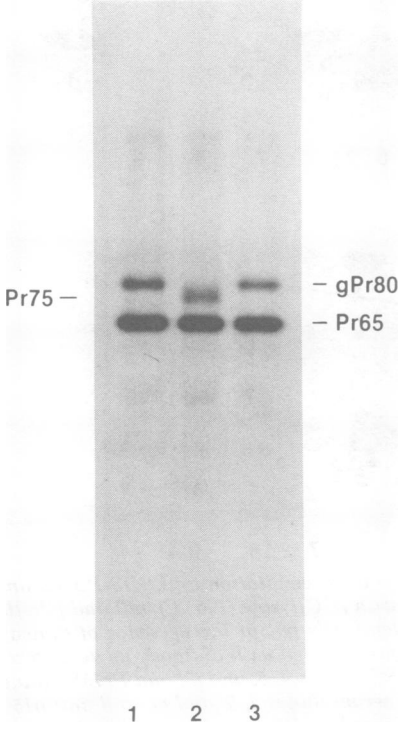

FIG. 1. *SDS-PAGE analysis of the effect of endoglycosidase digestion of R-MuLV gag polyproteins. JLS-V9 cells were labeled with [^{35}S]methionine (10 μCi/ml) for 15 min, lysed, and immunoprecipitated with anti-p30 goat serum. Immunoprecipitates were resuspended in 0.2 M sodium acetate buffer (pH 5.0) and treated for 4 h at 31°C with no enzyme (lane 1), 0.01 U of endoglycosidase H (lane 2), or 0.01 U of endoglycosidase D (lane 3).*

The substantial amount of uncleaved polyprotein present indicates the mildness of the hydrolysis conditions. The fragments identified with anti-p15 serum (lanes 3, 6, and 9) are those from the amino-terminal region of the polyproteins. The largest amino-terminal fragments (fragment A, 33K; fragment B, 22K; and fragment C, 29K) are different in size and represent the entire region of each polyprotein from its respective amino terminus to the common acid-sensitive site in p12. Partial acid cleavage occurs at other sites within those fragments. The smallest fragment identified with anti-p15 serum (fragment d, 10K), is common to all three proteins. Fragments a (23K), b (12K), and c (19K) from gPr80^{gag}, Pr65^{gag}, and Pr75^{gag}, respectively, all differ in molecular weight and are considered to comprise, along with the common fragment d, the entire fragment A, B, or C. That is, fragment A of gPr80^{gag} is composed of fragments a plus d, fragment B of Pr65^{gag} is composed of fragments b plus d, and fragment C of Pr75^{gag} is composed of fragments c plus d. The molecular weight estimates for fragments A, B, C and a, b, c agree well with this interpretation (see Table 1).

The large amino-terminal fragments of gPr80^{gag} (fragment A, 33K) and Pr65^{gag} (frag-

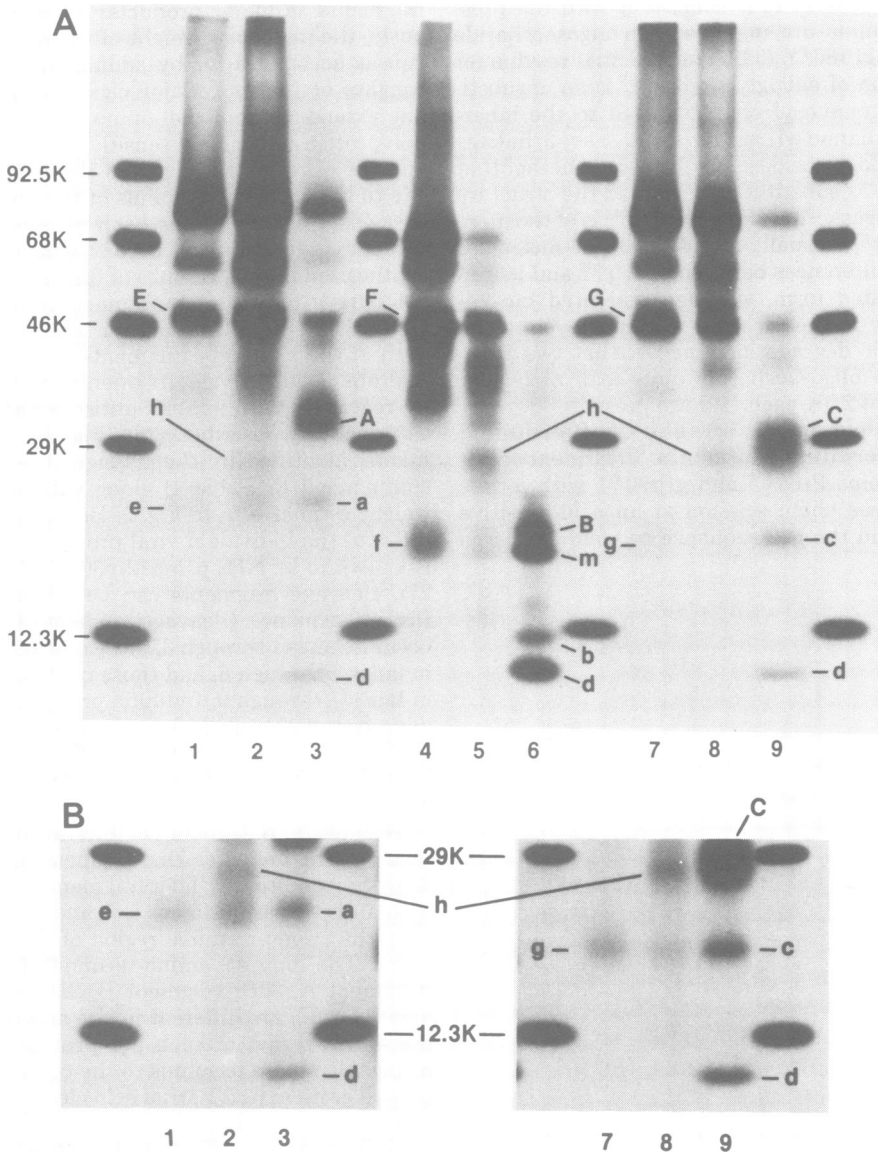

FIG. 2. *(A) Acid cleavage fragments of gag polyproteins: immunoprecipitation and SDS-PAGE analysis. Pr65^{gag} and gPr80^{gag} were isolated from JLS-V9 cells labeled with [^{14}C]lysine (125 μCi/ml) and [^{14}C]leucine (125 μCi/ml), and Pr75^{gag} was obtained from cells labeled identically but in the presence of tunicamycin. Polyproteins were purified, and fragments were obtained and precipitated with antibody as described in the text. Cleavage fragments from gPr80^{gag} (lanes 1, 2, and 3), Pr65^{gag} (lanes 4, 5, and 6), and Pr75^{gag} (lanes 7, 8, and 9) precipitated with anti-p10 serum (1, 4, and 7), anti-p30 serum (lanes 2, 5, and 8), and anti-p15 serum (lanes 3, 6, and 9) are shown. Molecular weight markers in unnumbered lanes are phosphorylase b (92.5K), bovine serum albumin (69K), ovalbumin (46K), carbonic anhydrase (29K), and cytochrome c (12.3K). (B) The low-molecular-weight fragments of gPr80^{gag} (lanes 1, 2, and 3) and Pr75^{gag} (lanes 7, 8, and 9) from (A) were reproduced on low-contrast photographic paper for visual clarity.*

ment B, 22K) differ in size by approximately 11K, in agreement with our previous finding (28). However, it is now evident that fragment A contains at least one carbohydrate unit in addition to the leader peptide because the analogous fragment from the unglycosylated Pr75^{gag} (fragment C, 29K) has an apparent molecular weight that is 4K smaller than that of fragment A from gPr80^{gag}.

Fragments A, B, and C exhibit characteristic

TABLE 1. *Acid cleavage fragments of gag polyproteins*

gPr80^{gag}		Pr65^{gag}		Pr75^{gag}		Reactive with antibody to:
Fragment	M_r ($\times 10^3$)	Fragment	M_r ($\times 10^3$)	Fragment	M_r ($\times 10^3$)	
A	33	B	22	C	29	p15
a	23	b	12	c	19	p15
d	10	d	10	d	10	p15
E	50	F	47	G	47	p10, p30
e	22.5	f	18.5	g	18.5	p10, p30
h	28	h	28	h	28	p30

molecular weight differences depending on the presence of the leader peptide or leader peptide plus carbohydrate. The same differences also occur in the further cleavage fragments a, b, and c. These smaller fragments (a, b, and c) thus represent the amino termini of the respective fragments A, B, and C. The other product of the further cleavage of fragments A, B, and C is the common fragment d, representing the carboxyl-terminal peptides of A, B, and C. The size and location of fragment d define that region of glycosylated fragment A which is free of carbohydrate. That is, the C-terminal 10K portion (i.e., fragment d) of fragment A, B, or C does not contain a glycosylation site, thus eliminating the p12 sequence of fragment A as a possible location of carbohydrate and restricting the position of the glycosylation site(s) to the leader peptide itself or to the amino-terminal three-fourths of the p15 sequence. (See Fig. 6 for a schematic representation.)

The leader peptide molecular weight can be estimated by comparing fragments of Pr65^{gag} with those of Pr75^{gag}, and the contribution of carbohydrate can be estimated by comparing fragments of Pr75^{gag} with those of gPr80^{gag}. The results are the same whether one compares fragments A, B, and C or fragments a, b, and c. That is, the 7K size difference between fragments B and C (29K and 22K) and also between fragments b and c (19K and 12K) defines the apparent molecular weight of the leader peptide. Similarly, the 4K-dalton contribution of carbohydrate to the apparent molecular weight can be obtained by comparing fragments A and B (33K and 29K) or fragments a and b (23K and 19K).

Fragment m (lane 6), precipitated by anti-p15 serum, is slightly smaller than fragment B and is thus a cleavage product of B. If m were produced from fragment B by removal of a short amino-terminal peptide, that same cleavage would also occur at the same sequence of Pr75^{gag} and fragment m would also be found in lane 9. Since that is not the case, fragment m must extend to the N-terminal of Pr65^{gag} and be produced by removal of a short carboxyl-terminal peptide from fragment B. Its counterpart in Pr75^{gag} and gPr80^{gag} will thus be significantly larger, the Pr75^{gag} counterpart containing the leader peptide and the gPr80^{gag} counterpart containing the leader peptide plus carbohydrate. The small difference in size between fragments B and m implies that the counterpart of fragment m from Pr75^{gag} will be similar in size to fragment C, and the counterpart of fragment m from gPr80^{gag} will be similar in size to fragment A. The broadness of the radioactive bands labeled fragments A and C in Fig. 2 is caused by the overlapping electrophoretic mobility of fragment A with the counterpart of fragment m from gPr80^{gag} and of fragment C with the counterpart of fragment m from Pr75^{gag}. In some experiments (not shown), the bands labeled A and C were resolved more clearly into doublets.

Acid cleavage: carboxyl-terminal fragments. The same cleavage event which generated amino-terminal fragments A, B, and C from gPr80^{gag}, Pr65^{gag}, and Pr75^{gag}, respectively, simultaneously produces a corresponding piece from each representing the carboxyl portions of the molecules. These fragments (fragments E, F, and G) react with anti-p10 and anti-p30 sera (lanes 1 and 2, 4 and 5, 7 and 8). The reactivity of these fragments with anti-p15 serum (lanes 3, 6, and 9) is attributable to anti-p12 activity in that serum. A close examination of band positions in Fig. 2 shows that fragments F and G have an identical migration, similar to that of the ovalbumin marker, whereas fragment E has a slightly slower mobility. That is, Pr65^{gag} and Pr75^{gag} contain fragments from the p30 to p10 portion of the molecule which are indistinguishable on the basis of size, whereas the fragment from the identical region of the glycosylated form, gPr80^{gag}, is larger. In the gradient gel of Fig. 2, the absolute mobility difference between E and fragments F and G represents an apparent molecular weight difference of 3K.

The utility of immune precipitation for identification of fragments is further emphasized by comparing the smallest peptides (e, f, and g) reactive with anti-p10 and anti-p30 sera. Such peptides are recovered in low yield from Pr75^{gag} and gPr80^{gag} but, nonetheless, the interpretation is clear. Fragment g from Pr75^{gag} is detectable with both anti-p10 and anti-p30 sera (lanes 7

and 8) and has the same molecular weight (18.5K) as fragment f from Pr65^{gag} (lanes 4 and 5). Neither fragment should be confused with fragments B and m (lane 6) or c (lane 9), even though the molecular weights are very similar, because of the specific reactions with the different antisera. Fragments f and g, from the p30 to p10 region of Pr65^{gag} and Pr75^{gag}, respectively, are identical in size, but the smallest peptide reactive with anti-p10 and anti-p30 sera from gPr80^{gag} (fragment e, 22.5K; lane 1) is larger. A fragment reactive with anti-p30 but not with anti-p10 serum (fragment h, lanes 2, 5, and 8) has a molecular weight of 28K and is common to all three polyproteins. If the molecular weight of fragment h is added to that of fragment e, f, or g, the sum is the molecular weight of fragment E, F, or G, respectively (see Table 1). Fragment h thus represents the amino-terminal region of fragments E, F, and G lacking p10 determinants, after fragment e, f, or g has been split off. Fragments e, f, and g, which are the carboxyl-terminal regions of fragments E, F, and G, respectively, obviously contain a carbohydrate attachment site sequence because of the size difference (4K) between fragment e (22.5K) from gPr80^{gag} and fragments f and g (18.5K) from the unglycosylated polyproteins. Further evidence for the precise location of the carbohydrate of fragment e will be presented below.

Fortuitously, the acid cleavage glycopeptides of gPr80^{gag} identified by specific antisera as originating from the amino-terminal (fragment a) and carboxyl-terminal (fragment e) regions have nearly identical molecular weights. The same is true for the corresponding unglycosylated peptides from Pr75^{gag} (fragments g and c). The contribution of carbohydrate to the apparent molecular weights of each glycosylated fragment, found by subtracting the apparent molecular weights of the respective unglycosylated peptide, is the same for each glycopeptide, namely, 4K. Since these molecular weight estimates are made from the same region of the polyacrylamide gel, it is very likely that the contribution of carbohydrate to the apparent molecular weight of both the amino- and carboxy-terminal glycopeptides is identical and that each peptide contains the same number of oligosaccharide chains.

Tryptic digestion of *gag* proteins. Experiments such as that shown in Fig. 2 examined the entire length of the *gag* polyproteins for the presence of carbohydrate by labeling with common amino acids (viz., leucine and lysine) and generating large fragments. At least two carbohydrate attachment sites were found in this way, but their location is very imprecise. An alternative approach is to label with a rare amino acid (e.g., methionine) and examine smaller portions of the molecule after more extensive fragmentation. Based on amino acid compositions of p15, p12, p30, and p10 (25), there are only three methionine residues in the entire Pr65^{gag}, therefore, the chances of finding a carbohydrate attachment site in this way are small. Fortunately, however, we were able to locate one carbohydrate site near a methionine residue (Fig. 3).

All three *gag* polyproteins and p30 labeled with [^{35}S]methionine were digested with trypsin and electrophoresed on slab polyacrylamide gels, using the original disc gel formulation of Davis (7). The radioautogram appears in Fig. 3.

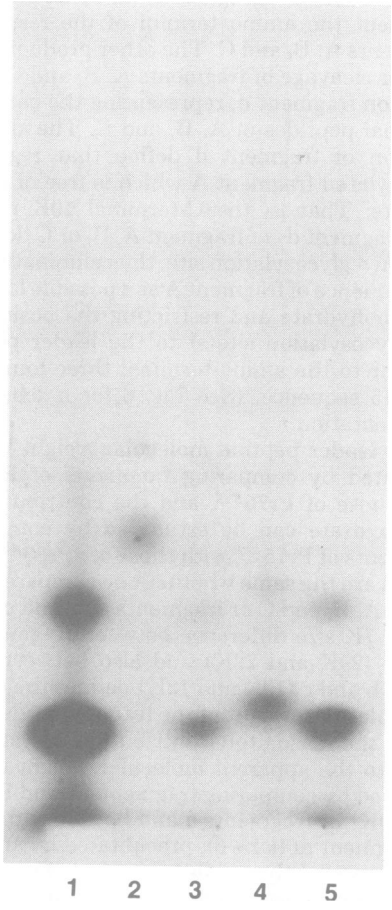

FIG. 3. *One-dimensional tryptic peptide maps of fragments of [^{35}S]methionine-labeled gag gene products. Proteins purified by immunoprecipitation and SDS-PAGE were trypsinized and electrophoresed under nondenaturing conditions (8) as described in the text. [^{35}S]methionine tryptic peptides from the following proteins are analyzed: p30 (lane 1); gPr80^{gag} (lane 2); Pr65^{gag} (lane 3); Pr75^{gag} generated by endoglycosidase H digestion (lane 4); Pr75^{gag} synthesized in the presence of tunicamycin (lane 5).*

The radioactivity moving with the tracking dye (bottom) is due to a very small tryptic peptide(s). The majority of [^{35}S]methionine-labeled material is resolved by the gel into a major and a minor spot produced from p30 (lane 1). The same pair of peptides is obtained from Pr65^{gag} (lane 3) and Pr75^{gag} synthesized in the presence of tunicamycin (lane 5). Thus, the methionine-containing peptides of Pr65^{gag} and Pr75^{gag} which appear in Fig. 3 originate in the p30 region of the polyproteins. However, these p30 peptides are not produced from gPr80^{gag} (lane 2). Instead, there appears a different pair of peptides with considerably less mobility. In this gel system, which contains no SDS, peptides migrate according to the intrinsic charge/mass ratio they possess at pH 8.3. Since gPr80^{gag} and Pr75^{gag} are the same protein, one with carbohydrate and one without, they share the same tryptic cleavage sites. What is responsible for the altered mobility of the p30 peptides of gPr80^{gag}?

A straightforward interpretation of the gel in Fig. 3 is that a p30 tryptic peptide containing methionine can be found in gPr80^{gag} as a glycopeptide carrying the mannose-rich core oligosaccharide. The intrinsic charge of the apopeptides from p30, Pr65^{gag}, Pr75^{gag}, and gPr80^{gag} are all the same since they represent the identical amino acid sequence. The only sugars found in mannose-rich oligosaccharides are neutral sugars (19); they carry no electrical charge. Therefore, the glycopeptide of gPr80^{gag} has substantially greater mass and unaltered charge; thus, its mobility in the electrical field is reduced.

In this context, it is interesting to observe lane 4, which shows the tryptic peptides obtained from Pr75^{gag} generated not with tunicamycin, as in lane 5, but by endoglycosidase H digestion of gPr80^{gag} and subsequent trypsinization. The mobility of the pair of p30 peptides from this molecule is slightly less than that of Pr75^{gag} produced by tunicamycin (lane 5). The bond cleaved by endoglycosidase H is that which links the two *N*-acetylglucosaminyl residues at the reducing end of the mannose-rich oligosaccharide (33). The product of digestion is a glycopeptide with one *N*-acetylglucosaminyl residue still attached to the asparagine residue of the polypeptide chain. Thus, the tryptic peptide generated from tunicamycin-produced Pr75^{gag}, which is devoid of asparagine-linked carbohydrate, is not precisely the same as that produced after endoglycosidase H digestion. The additional 225 daltons of mass due to the remaining *N*-acetylglucosaminyl moiety increases the mass sufficiently to produce the slower mobility seen in lane 4.

Tryptic peptide containing a glycosylation site. The complete sequence of p30 is known from work in our laboratory (24a; L. E. Henderson et al., unpublished data). The amino acid sequence of a region of p30 which contains methionine in which the amino-terminal tyrosine residue has been generated by tryptic cleavage is shown in Fig. 4. The peptide contains three aspartic acid and two glutamic acid residues for a total of five acidic residues and one or two basic residues, depending on whether trypsin cleaves at the arginine or lysine residue to generate the carboxyl-terminus. If trypsin cleaves at the arginine residue, the peptide will be 29 amino acids long. If the initial tryptic cleavage occurs at the lysine residue, subsequent cleavage to yield the 29-residue peptide will be relatively slow because trypsin is inefficient as an exopeptidase. Since the peptide contains five acidic residues, the net charge will be −4 or −3 depending on the actual cleavage site. The absence or presence of the additional lysine residue has only a small effect on the peptide mass, but the effect on charge is substantial. Thus, the major and minor spots of Fig. 3 correspond to the two possible tryptic cleavage sites for the single sequence, the major spot being the cleavage product (arginine carboxyl terminus) with a net charge of −4 and the minor spot being the 30-residue (lysine carboxyl terminus) peptide with a −3 charge and a slower mobility.

That the major peptide of Fig. 3 (lane 1) represents the sequence shown in Fig. 4 has been confirmed by stepwise Edman degradation of a [^{35}S]methionine-labeled peptide from p30 isolated from a gel identical to Fig. 3. The detection of ^{35}S-labeled radionuclide at each step in the sequential degradation is given in Fig. 5. A single

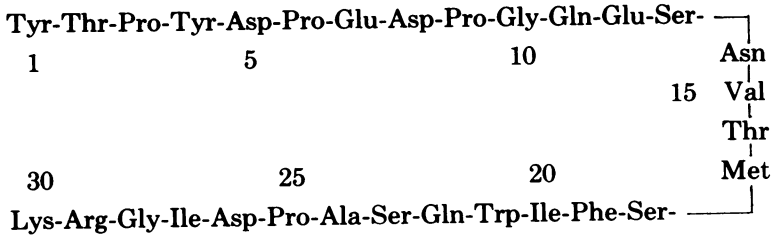

FIG. 4. *Cleavage map of R-MuLV gag polyproteins.*

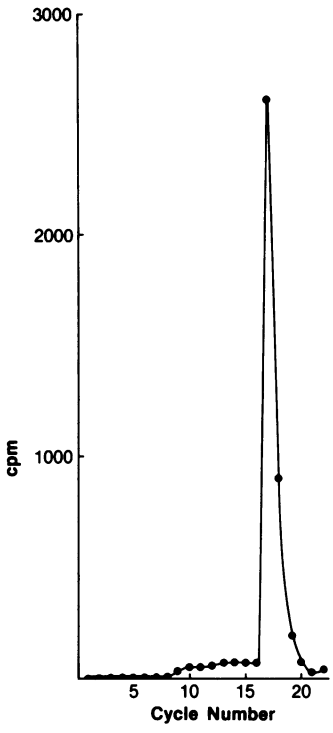

FIG. 5. *Sequence of a methionine-containing tryptic peptide from p30.*

peak of radioactivity was detected at step 17, exactly the position in which methonine occurs in the amino acid sequence shown in Fig. 4. In data not shown, tyrosine, proline, and valine occurred in positions consistent with this sequence when a tryptic peptide labeled with [^{3}H]tyrosine, [^{3}H]proline, and [^{3}H]valine, isolated in the same way, was analyzed.

Empirical observation suggests that asparagine residues may bear oligosaccharide chains only if the residue on the carboxyl side of the asparagine residue is followed by a serine or threonine residue (21). The peptide shown in Fig. 4 contains such a tripeptide, the Asn-Val-Thr sequence immediately preceding the methionine residue at position 17. Our experiments define this asparagine residue as a glycosylation site of R-MuLV gPr80^{gag}.

The acid cleavage site responsible for generation of the smallest peptides (e, f, and g) reactive with anti-p30 and anti-p10 sera is also found in the peptide shown in Fig. 4. Aspartic acid residues followed by proline residues are especially susceptible to acid-catalyzed cleavage through formation of a five-membered ring intermediate (26). Two such Asp-Pro sequences, the only ones in the entire R-MuLV p30 (Henderson et al., unpublished data), occur in the first nine residues shown in Fig. 4. The tyrosine at position 1 is known from sequence analysis to be 100 residues away from the carboxyl terminus of p30. Cleavage at either Asp-Pro linkage thus produces from p30 a peptide of approximately 11K in size. The molecular weight of p10, determined from the complete primary structure (L. E. Henderson, T. D. Copeland, R. C. Sowder, G. W. Smythers, and S. Oroszlan, J. Biol. Chem., in press), is 6.347K. We can predict, therefore, that the largest possible fragment originating from acid cleavage of Pr65^{gag} or Pr75^{gag} at an Asp-Pro bond and reactive with anti-p30 and anti-p10 sera will have a size of approximately 17K. Fragments f and g (Fig. 2) from Pr65^{gag} and Pr75^{gag}, respectively, are reactive with anti-p10 and anti-p30 sera and are at least as large (18.5K ± 10%) as the predicted size. Fragments f and g originating in this way contain the Asn-Val-Thr tripeptide responsible for the carbohydrate found in the corresponding fragment e from gPr80^{gag}.

DISCUSSION

The *gag* polyproteins of MuLV are interesting because they are distributed between two biosynthetic pathways, with one (Pr65^{gag}) leading to viral structural proteins and another (gPr80^{gag}) leading to glycosylated cell surface components which are apparently not present in the virion. Previously, we found that one structural difference between gPr80^{gag} and Pr65^{gag} was the presence of a long leader peptide. This raised the possibility that this leader could function as a signal sequence, thus accounting for the different posttranslational processing pathways of two otherwise virtually identical gene products (1, 13). In this study we have obtained an estimate for the size of the leader peptide, shown that gPr80^{gag} contains only two immature mannose-rich oligosaccharides, and specified their locations.

The results which support these conclusions are summarized in a physical map of the polyproteins (Fig. 6). Acid cleavage fragments from Pr65^{gag}, Pr75^{gag}, and gPr80^{gag} are aligned on the map with molecular weights from Table 1. The location of carbohydrate is indicated by the schematic branched structures attached to gPr80^{gag}. Also included in Fig. 6, for the purposes of comparison, is the carboxyl-terminal cyanogen bromide fragment of Pr65^{gag} and gPr80^{gag} from a previous study (28). The amino-terminal fragment m (18K) of Pr65^{gag} and the corresponding fragments from Pr75^{gag} (25K) and gPr80^{gag} (29K) are also indicated.

A comparison of fragment sizes between Pr65^{gag} and Pr75^{gag} shows that all are identical

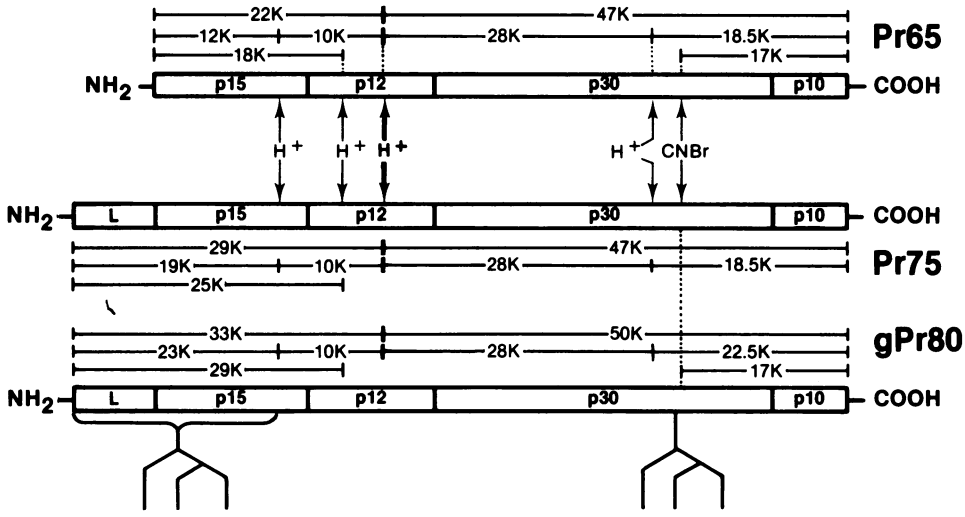

FIG. 6. *Stepwise degradation of a [^{35}S]methionine-labeled tryptic peptide. The major tryptic peptide from p30 was isolated from a gel like that shown in Fig. 3 and subjected to a stepwise Edman degradation as described in the text. Fractions recovered at each cycle were dried and analyzed for the presence of ^{35}S-labeled radionuclide by scintillation counting.*

except for those containing the extreme amino-terminal sequences. This result confirms that Pr75^{gag} bears a leader peptide whose size can be estimated from this data to be approximately 7K. Similar experiments with the *gag* precursor from M-MuLV utilizing cleavage with V8 protease (9) or hydroxlamine (C. J. M. Saris, H. C. M. Van Eenbergen, R. Liskamp, and H. P. J. Bloemers, personal communication) have given values of 4K and 6K, respectively, for the leader peptide of that viral gene product. Although it is possible that the extra sequences are inserted somewhere within the amino-terminal region, we prefer the more likely possibility that the extra sequences exist as a leader peptide attached to the amino terminus (28).

The total carbohydrate value obtained by comparing the SDS-PAGE mobilities of Pr75^{gag} (75K) and gPr80^{gag} (80K) is 5K, a size difference consistent with two mannose-rich oligosaccharides of 2.5K each. However, as smaller fragments containing carbohydrate are examined, the molecular weight contribution of carbohydrate based on SDS-PAGE mobilities increases to 4K in each glycopeptide, raising the possibility of more than one oligosaccharide at each peptide. There are, however, two and only two oligosaccharides in the entire gPr80^{gag}, and the argument is as follows. The central part of the polyproteins is devoid of carbohydrate, as indicated by fragments d and h (10K and 28K, respectively), which are common to glycosylated and unglycosylated polyproteins alike. Glycosylation is therefore limited to the extreme amino- and carboxyl-terminal fragments. At the carboxyl end, fragments f and g from Pr65^{gag} and Pr75^{gag}, respectively (18.5K), contain the sequences which are glycosylated in fragment e (22.5K) of gPr80^{gag}. From other work in our laboratory, the complete primary structure of p10 (Henderson et al., J. Biol. Chem., in press) and p30 (Henderson et al., unpublished data) is established. Therefore, we know that, except for the glycosylation site shown in Fig. 4, no other Asn-X-Ser/Thr tripeptide sequences required for glycosylation are present in the sequence of fragment e, and consequently, this fragment can contain only one oligosaccharide unit. The molecular weight contribution allotted to carbohydrate in fragment e (22.5K) of gPr80^{gag}, found by subtracting the molecular weight of corresponding fragment f (18.5K) of Pr75^{gag}, is 4K. Therefore, the single oligosaccharide in fragment e increases the SDS-PAGE molecular weight estimate by 4K. On the other hand, the smallest glycosylated fragment (23K) from the amino-terminal region of gPr80^{gag}, which is nearly identical in size to fragment e, also has 4K attributed to carbohydrate by comparison with the corresponding fragment from Pr75^{gag} (19K). This makes a strong case for the presence of a single oligosaccharide in the amino-terminal region also. We therefore conclude that the carbohydrate of gPr80^{gag} consists of two mannose-rich core oligosaccharides, one found in a 23K fragment located in the amino-terminal region of gPr80^{gag} and containing the additional amino acid sequences not found in Pr65^{gag} plus a sub-

stantial portion of p15. The second site is clearly shown to occur within p30 at an asparagine residue, which from other work (Henderson et al., unpublished data) is known to occur in position 177 of that protein sequence.

Similar experiments on *gag* polyproteins of Mo-MuLV have been performed by Edwards and Fan (9), using V8 proteolytic cleavage. They concluded that a common amino-terminal region of the polyproteins extending 5K into the p30 region (32K polypeptide from the amino-terminus in $Pr65^{gag}$), contains the glycosylation sites. In R-MuLV the glycosylation site we have structurally defined in the p30 region (Asn_{177}) is outside and C-terminal to this segment. Protein sequence analysis in our laboratory and nucleic acid sequence data (C. Van Beveren and I. M. Verma, personal communication) indicate a homologous sequence around Asn_{177} which is a potential glycosylation site also in Mo-MuLV p30. It is possible that the actual glycosylation site in R-MuLV and Mo-MuLV are different. However, the exact location of the carbohydrate in the p30 region of Mo-MuLV must await structural analysis as done here for R-MuLV. It is possible that the two viruses share a glycosylation site in the amino-terminal region, since the carbohydrate-containing region of Mo-MuLV proposed by Edwards and Fan extends into the p15, slightly overlapping our proposed location for carbohydrate in R-MuLV.

It is now possible to reinterpret the results of previous analyses of tryptic digests of $Pr65^{gag}$ and $gPr80^{gag}$. Kopchik et al. (14), utilizing a [^{14}C]tyrosine radiolabel, reported that two tryptic peptides of R-MuLV $Pr65^{gag}$ were not found in $gPr80^{gag}$. The tryptic peptide shown in Fig. 4 contains two tyrosine residues and would, therefore, be prominent in a digest of [^{14}C]tyrosine-labeled protein. From Fig. 3, it is likely that this peptide sequence would yield two tryptic peptides from $Pr65^{gag}$. These would not appear in an analysis of $gPr80^{gag}$ because the corresponding glycopeptides would probably pass unretarded through the ion-exchange column used for separating the peptides. Therefore, the lack of these $Pr65^{gag}$ peptides in $gPr80^{gag}$ is not caused by sequence differences but is adequately explained by glycosylation of the tryptic peptide of Fig. 4. This is similar to the findings and suggestion of Ledbetter et al. (17), who observed differences between peptide maps of glycosylated and unglycosylated *gag* polyproteins of Gross MuLV.

Previously, we have presented two-dimensional tryptic maps of [^{35}S]methionine-labeled $Pr65^{gag}$ and $gPr80^{gag}$ which were identical with each other and with maps of p30 (28). At that time, it was known that the methionine residue of p12 was not represented in the maps, and it was suggested that the two radioactive p30 spots represented the two methionine residues of p30. It is now clear that the 29- or 30-residue tryptic peptide(s) of p30 (Fig. 4) probably remained at the origin of these maps and that the two tryptic map spots observed represent overlapping tryptic peptides containing the other methionine residue of p30 found with the tracking dye (see Fig. 3).

The methionine residue of the peptide shown in Fig. 4 (residue 180 in the p30 sequence) is the cleavage site for cyanogen bromide resulting in the 17K carboxy-terminal fragment common to $Pr65^{gag}$ and $gPr80^{gag}$ (28), as indicated in Fig. 6. The present results demonstrate that these previously observed cyanogen bromide fragments lack the glycosylation site for the single oligosaccharide unit occurring in the carboxyl-terminal region.

According to the signal hypothesis (4), amino acid sequences preceded by a signal peptide are cotranslationally transported across the membrane of the endoplasmic reticulum and glycosylated at appropriate sites. Secretory glycoproteins pass completely through the membrane into the lumen, whereas transmembrane proteins contain hydrophobic sequences which remain associated with the membrane. As shown in the present study, R-MuLV $gPr80^{gag}$ is glycosylated at Asn_{177} of the p30 sequence. Beyond that point in the sequence, we know that the entire p10 and the carboxyl third of p30 are extremely polar in nature (Henderson et al., J. Biol. Chem., in press; Henderson et al., unpublished data). If, according to the signal hypothesis, the entire $gPr80^{gag}$ sequence up to Asn_{177} of p30 has been transported into the lumen of the endoplasmic reticulum, there are thus no hydrophobic domains in the subsequently synthesized sequence for $gPr80^{gag}$ to become anchored in the membrane. After it passes into the Golgi membranes, where immature carbohydrate structures are completed, the lack of membrane attachment would account for its subsequent export from the cell as a secretory glycoprotein, $gP94^{gag}$ (11, 17, 18, 29).

Although it has not been fully proven, it is very likely that the $Pr65^{gag}$ sequences in $Pr65^{gag}$ and $Pr75^{gag}$ are essentially identical, and that the glycosylation of $Pr75^{gag}$ at sites which are not glycosylated in $Pr65^{gag}$ occurs because a signal peptide attached to the $Pr75^{gag}$ sequence permits association with glycosylating membranes. Whether the leader peptide itself contains a signal sequence is not yet known; the existence of a separate signal peptide which is cotranslationally cleaved from $gPr80^{gag}$ is not excluded, although it has been observed that

$Pr75^{gag}$ synthesized in vivo and in vitro does not differ in size (9). It is noteworthy that the leader peptide of R-MuLV *gag* polyproteins is significantly larger than the usual 15- to 30-residue signal sequences. The roles of the leader peptide and of the *gag* polyproteins themselves remain to be elucidated.

ACKNOWLEDGMENTS

We thank Robert Hamill for the gift of tunicamycin and Terry Copeland for sequence analysis of radioactive peptides.

This work was supported by Public Health Service contract N01-CO-75380 from the National Cancer Institute.

LITERATURE CITED

1. **Arcement, L. J., W. L. Karshin, R. B. Naso, and R. B. Arlinghaus.** 1977. *gag* polyprotein precursors of Rauscher murine leukemia virus. Virology **81**:284–297.
2. **Arcement, L. J., W. L. Karshin, R. B. Naso, G. Jamjoom, and R. B. Arlinghaus.** 1976. Biosynthesis of Rauscher leukemia viral proteins: presence of p30 and envelope p15 sequences in precursor polypeptides. Virology **69**:763–774.
3. **Barbacid, M., J. R. Stephenson, and S. A. Aaronson.** 1976. *gag* gene of mammalian type-C RNA tumor viruses. Nature (London) **262**:554–559.
4. **Blobel, G., and B. Dobberstein.** 1975. Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. J. Cell Biol. **67**:835–851.
5. **Butterworth, B. E.** 1977. Proteolytic processing of animal virus proteins. Curr. Top. Microbiol. Immunol. **77**:1–41.
6. **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.
7. **Davis, B. J.** 1964. Disc electrophoresis—II, method and application to human serum proteins. Ann. N.Y. Acad. Sci. **121**:404–427.
8. **Edwards, S. A., and H. Fan.** 1979. *gag*-related polyproteins of Moloney murine leukemia virus: evidence for independent synthesis of glycosylated and unglycosylated forms. J. Virol. **30**:551–563.
9. **Edwards, S. A., and H. Fan.** 1980. Sequence relationship of glycosylated and unglycosylated *gag* polyproteins of Moloney murine leukemia virus. J. Virol. **35**:41–51.
10. **Eisenman, R. N., and V. M. Vogt.** 1978. The biosynthesis of oncovirus proteins. Biochim. Biophys. Acta **473**:187–239.
11. **Evans, L. H., S. Dresler, and D. Kabat.** 1977. Synthesis and glycosylation of polyprotein precursors to the internal core proteins of Friend murine leukemia virus. J. Virol. **24**:865–874.
12. **Hunkapiller, M. W., and L. E. Hood.** 1978. Direct microsequence analysis of polypeptides using an improved sequenator, a nonprotein carrier (polybrene), and high pressure liquid chromatography. Biochemistry **17**:2124–2133.
13. **Jamjoom, G. A., R. B. Naso, and R. B. Arlinghaus.** 1977. Further characterization of intracellular precursor polyproteins of Rauscher leukemia virus. Virology **78**: 11–34.
14. **Kopchick, J. J., W. L. Karshin, and R. B. Arlinghaus.** 1979. Tryptic peptide analysis of *gag* and *gag-pol* gene products of Rauscher murine leukemia virus. J. Virol. **30**:610–623.
15. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
16. **Laskey, R. A., and A. D. Mills.** 1975. Quantitative film detection of ^{3}H and ^{14}C in polyacrylamide gels by fluorography. Eur. J. Biochem. **56**:335–341.
17. **Ledbetter, J., R. C. Nowinski, and R. Eisenman.** 1978. Biosynthesis and metabolism of viral protein expression on the surface of murine leukemia virus-infected cells. Virology **91**:116–129.
18. **Ledbetter, J., R. C. Nowinski, and S. Emery.** 1977. Viral proteins expressed on the surface of murine leukemia cells. J. Virol. **22**:65–73.
19. **Li, E., I. Tabas, and S. Kornfeld.** 1978. The synthesis of complex-type oligosaccharides. I. Structure of the lipid-linked oligosaccharide precursor of the complex-type oligosaccharides of the vesicular stomatitis virus G protein. J. Biol. Chem. **253**:7762–7770.
20. **Long, C. W., R. Berzinski, and R. V. Gilden.** 1977. Immunologic studies of the low molecular weight DNA binding protein of murine oncornaviruses. Int. J. Cancer **19**:843–850.
21. **Marshall, R. D.** 1974. The nature and metabolism of carbohydrate-peptide linkage of glycoproteins. Biochem. Soc. Symp. **40**:17–26.
22. **Old, L. J., E. A. Boyse, and E. Stockert.** 1965. The G (Gross) leukemia antigen. Cancer Res. **25**:813–819.
23. **Oroszlan, S., T. D. Copeland, L. E. Henderson, J. R. Stephenson, and R. V. Gilden.** 1979. Amino-terminal sequence of bovine leukemia virus major internal proteins: homology with mammalian type C virus p30 structural proteins. Proc. Natl. Acad. Sci. U.S.A. **76**: 2996–3000.
24. **Oroszlan, S., C. L. Fisher, T. B. Stanley, and R. V. Gilden.** 1970. Proteins of the murine C-type RNA tumour viruses: isolation of group-specific antigen by isoelectric focusing. J. Gen. Virol. **8**:1–10.

24a. **Oroszlan, S., L. E. Henderson, T. D. Copeland, A. M. Schultz, and E. M. Rabin.** 1980. Processing and structure of murine leukemia virus *gag* and *env* gene encoded polyproteins, p. 219–232. *In* G. Koch and D. Richter (ed.), Biosynthesis, modification and processing of cellular and viral polyproteins. Academic Press, Inc., New York.

25. **Oroszlan, S., L. E. Henderson, J. R. Stephenson, T. D. Copeland, C. W. Long, J. N. Ihle, and R. V. Gilden.** 1978. Amino- and carboxyl-terminal amino acid sequences of proteins coded by *gag* gene of murine leukemia virus. Proc. Natl. Acad. Sci. U.S.A. **75**:1404–1408.
26. **Piszkiewicz, D., M. Landon, and E. L. Smith.** 1970. Anomalous cleavage of aspartyl-proline peptide bonds during amino acid sequence determinations. Biochem. Biophy. Res. Commun. **40**:1173–1178.
27. **Reynolds, R. K., and J. R. Stephenson.** 1977. Intracistronic mapping of the murine type C viral *gag* gene by use of conditional lethal replication mutants. Virology **81**:328–340.
28. **Schultz, A. M., and S. Oroszlan.** 1978. Murine leukemia virus *gag* polyproteins: the peptide chain unique to Pr80 is located at the amino terminus. Virology **91**:481–486.
29. **Schultz, A. M., E. H. Rabin, and S. Oroszlan.** 1979. Post-translational modification of Rauscher leukemia virus precursor polyproteins encoded by the *gag* gene. J. Virol. **30**:255–266.
30. **Schultz, A. M., S. K. Ruscetti, E. M. Scolnick, and S. Oroszlan.** 1980. The *env*-gene of the spleen focus-forming virus lacks expression of p15(E) determinants. Virology **107**:537–542.
31. **Snyder, H. W., Jr., E. Stockert, and E. Fleissner.** 1977. Characterization of molecular species carrying Gross cell surface antigen. J. Virol. **23**:302–314.
32. **Tai, T., K. Yamashita, M. Ogata-Arakawa, N. Koide, T. Muramatsu, S. Iwashita, Y. Inoue, and A. Kobata.** 1975. Structural studies of two ovalbumin glycopeptides in relation to the endo-β-N-acetylglucosamin-

idase specificity. J. Biol. Chem. **250**:8569–8575.

33. **Tarentino, A. L., and F. Maley.** 1975. A comparison of the substrate specificities of endo-β-N-acetylglucosaminidases from *Streptomyces Griseus* and *Diplococcus Pneumoniae*. Biochem. Biophys. Res. Commun. **67**: 455–462.
34. **van Zaane, D., A. L. J. Gielkens, M. J. A. Dekker-Michielsen, and H. P. J. Bloemers.** 1975. Virus-specific precursor polypeptides in cells infected with Rauscher leukemia virus. Virology **67**:544–552.
35. **Vogt, V. M., and R. Eisenman.** 1973. Identification of a large polypeptide precursor of avian oncornavirus proteins. Proc. Natl. Acad. Sci. U.S.A. **70**:1734–1738.
36. **Wright, B. S., P. A. O'Brien, G. P. Shibley, S. A. Mayyasi, and J. C. Lasfargues.** 1967. Infection of an established mouse bone marrow cell line (JLS-V9) with Rauscher and Moloney murine leukemia viruses. Cancer Res. **27**:1672–1677.