JOURNAL OF VIROLOGY, Jan. 1981, p. 36-47
0022-538X/81/010036-12$02.00/0

Vol. 37, No. 1

Protein-Coding Potential of Mouse Mammary Tumor Virus Genome RNA as Examined by In Vitro Translation

CLIVE DICKSON* AND GORDON PETERS

Imperial Cancer Research Fund Laboratories, Lincoln's Inn Fields, London WC2A 3PX, United Kingdom

The protein-coding capacity of the mouse mammary tumor virus genome has been examined by in vitro translation of genome length and polyadenylated subgenomic fragments of viral RNA. Intact genome RNA of about 35S programmed synthesis of the $Pr77^{gag}$, $Pr110^{gag}$ and $Pr160^{gag/pol}$ precursors seen in infected cells in vivo. Polyadenylated RNA fragments of 18 to 28S encoded products whose tryptic peptide maps resembled those of the nonglycosylated precursor to the envelope glycoproteins, confirming the gene order 5'-*gag-pol-env*-3'. Translation of polyadenylated RNA fragments smaller than 18S yielded a series of related proteins whose peptide maps bore no resemblance to any of the virion structural proteins. Thus, a region of the mouse mammary tumor virus genome distal to the *env* gene appears to have an open reading frame sufficient to encode at least 36,000 daltons of protein as of yet unknown function.

Mouse mammary tumor virus (MuMTV) differs in several respects from other members of the retrovirus family. For example, it induces a high incidence of mammary adenocarcinomas, as opposed to the more common leukemias and sarcomas associated with other retroviruses; its expression is subject to hormonal control; and it is the prototype of the type B class of retroviruses as defined by morphological criteria (18, 27, 45, 46). Nevertheless, the available biochemical data imply that MuMTV has the same general strategy for replication as the other retroviruses (36, 43, 46).

The synthesis of infectious virus requires the expression of three viral genes (3). The *gag* gene encodes the internal structural proteins and, in the case of MuMTV, is expressed in the form of two polyprotein precursors; $Pr77^{gag}$ and $Pr110^{gag}$ (9, 10, 28, 34, 38). Previous studies have shown that these precursors can be synthesized in an in vitro translation system programmed by MuMTV genome RNA, suggesting that the *gag* gene is located at or near the 5' end of the genome (8, 28, 42). The *pol* gene encodes the virion RNA-dependent DNA polymerase or reverse transcriptase. In other systems, this region of the genome is expressed as a precursor which encompasses both the *gag* and *pol* genes (19, 22, 29, 33). In MuMTV, a Pr160 product which is detectable both in infected cells and by in vitro translation of viral RNA has been identified as the potential *gag-pol* precursor (8, 9). A third gene, *env*, encodes the glycoproteins of the viral envelope. In MuMTV, these proteins are synthesized as a $Pr73^{env}$ precursor which is subsequently processed to yield the gp52 and gp36 glycoproteins found in mature virus particles (12, 34, 38, 39, 42). By analogy with other systems, the MuMTV *env* gene is assumed to lie immediately distal to *pol*, making the gene order 5'-*gag-pol-env*-3' (3, 5, 48).

In some retroviruses, an additional gene has been identified which determines the oncogenic potential of the virus (4, 26). Characteristically, these viruses induce relatively rapid neoplasia in vivo and transform cells in vitro, and a high proportion of all infected cells become transformed (45). The additional information appears to be derived from a normal cellular element acquired either at the expense of some of the viral replicative functions, as is the case for the defective acute leukemia viruses, or in addition to the replicative functions, distal to the *env* gene, such as in nondefective avian sarcoma viruses (ASV) (5, 7, 30, 44; T. Graf, H. Beug, A. V. Kirchbach, and M. J. Hayman, Cold Spring Harbor Symp. Quant. Biol., in press). In ASV, these additional sequences code for a 60,000-dalton phosphoprotein with an associated protein kinase activity whose function is necessary for viral transformation (6, 25). This product is not normally found in mature virus particles but can be readily detected in ASV-transformed cells.

In contrast, no additional gene has yet been detected in a second major group of retroviruses, which includes the nondefective avian and murine leukosis viruses and possibly MuMTV. Although these viruses are oncogenic in vivo, the diseases induced arise after a long latency period and as a rare consequence of infection since characteristically many cells in the animal are

productively infected but appear phenotypically normal (27, 45; R. D. Cardiff and L. J. T. Young, *Viruses in Naturally Occurring Tumors*, in press). Such viruses frequently occur both as horizontally transmissible agents and as endogenous genetic elements (18, 45; Cardiff and Young, in press). Moreover, they do not transform cells in vitro so that it has not been possible to delineate a transforming function by selection of conditional mutants.

Although no functions other than *gag*, *pol*, and *env* have been identified in these viruses, including MuMTV, there is still a potential for the viral genome to encode additional, nonstructural proteins which may be implicated in oncogenicity. One approach to this question is to examine the protein-coding capacity of the viral RNA by in vitro translation (14, 31, 40). Retrovirus genome RNA is both capped and polyadenylated and of positive sense and can therefore serve as an mRNA for in vitro protein synthesis (16, 20, 47). This paper describes experiments designed to examine the total coding potential of the MuMTV genome in an attempt to identify possible nonstructural products and confirm the order of the structural genes.

MATERIALS AND METHODS

Cells and viruses. Two lines of mouse mammary carcinoma cells chronically releasing MuMTV were used as a source of virus. The first of these, derived from a mammary tumor in the GR mouse strain as described by Ringold et al. (35), was propagated in Dulbecco modified Eagle medium supplemented with 10% horse serum. The second line, Mm5MT/cl, was derived from a C3H mouse and grown in Eagle medium with 10% fetal calf serum (15). After initial seeding of the cells, cultures were maintained in medium containing 10 μg of dexamethasone per ml to stimulate virus production (15, 35). The production of MuMTV by these cell lines was routinely monitored by a reverse transcriptase assay with conditions appropriate for the distinction between types B and C murine retroviruses (50).

For large-scale preparation of virus, the cells were grown in roller culture on a Bellco automated harvesting device, with medium changes every 4 h. The cells were continuously labeled with [5-^{3}H]uridine (Radiochemical Centre; 25 Ci/mmol) at a concentration of 2 μCi/ml of medium. The pooled culture fluids were held at 0°C, and virus was prepared daily by high-speed centrifugation. Virus was further purified by equilibrium centrifugation at 240,000 × *g* for 1 h in linear gradients of 10 to 55% sucrose in TNE buffer (10 mM Tris-hydrochloride [pH 7.6], 100 mM NaCl, 1 mM EDTA). The virus band was recovered and stored at −70°C for up to 4 months.

Preparation of viral RNA. RNA was extracted directly from the pooled virus suspensions in sucrose. After an initial treatment of the suspension with self-digested pronase (500 μg/ml) for 10 min at room temperature, the virions were lysed by addition of sodium dodecyl sulfate to a final concentration of 0.5% (wt/vol). Digestion was continued for a further period of 10 min at room temperature, and the suspension was then extracted twice with an equal volume of TNE buffer-saturated phenol and once with chloroform-isoamyl alcohol (24:1). RNA was precipitated from the aqueous phase by addition of sodium acetate (0.2 M) and two volumes of ethanol at −20°C.

The viral RNA was fractionated into the 70 and 4S components by centrifugation through gradients of 5 to 20% sucrose in TNE buffer containing 0.5% Sarkosyl (Geigy). Centrifugation was performed at 240,000 × *g* for 75 min at 4°C. The fractions containing the 70S RNA were pooled and ethanol precipitated, and the recovery of RNA was assessed by measuring both the absorbance at 260 nm and the level of [^{3}H]uridine. The 70S RNA complex was then denatured by heating to 95°C for 2 min and applied directly to a 1-ml column of polyuridylic acid-Sepharose (13). The polyadenylated RNA, amounting to 25 to 35% of the total RNA applied, was eluted with distilled water and ethanol precipitated after addition of 50 μg of calf liver tRNA (Boehringer Mannheim) as carrier. This polyadenylated RNA was then further fractionated by sucrose gradient sedimentation at 240,000 × *g* for 2.5 h with ^{32}P-labeled 28 and 18S rRNA as internal markers. Samples of each fraction were used to determine the levels of ^{3}H and ^{32}P radioactivity by scintillation counting. The remainder of each fraction was ethanol precipitated with 10 μg of tRNA as the carrier and washed twice by reprecipitation to remove traces of detergent. Samples of the individual fractions were then used to program in vitro translation reactions as described below. In some experiments, the gradient fractions were pooled and precipitated as three major size classes: 35 to 28S, 28 to 18S, and 18 to 10S.

Limited degradation of viral RNA. Fractions containing the 35S polyadenylated RNA were recovered from sucrose gradients as described above. A portion of this RNA was treated with 50 mM Na_2CO_3 at 50°C as described by Coffin and Billeter (5). Samples were withdrawn into 100 mM Tris-hydrochloride (pH 7.6) after 30, 60, 90, and 120 s, combined, and then rechromatographed on polyuridylic acid-Sepharose. The polyadenylated RNA was recovered as before and, in some experiments, subjected to further size fractionation on sucrose gradients.

Cell-free protein synthesis. Rabbit reticulocyte lysates, capable of synthesizing protein in response to added mRNA, were prepared essentially as described by Pelham and Jackson (32) except that amino acids (minus methionine) were added to a concentration of 0.2 mM each, and 2-aminopurine was added to 6 mM. The reactions (20 μl) also contained 20 μCi of [^{35}S]-methionine (Radiochemical Centre) (600 to 1,100 Ci/mmol) and were initiated by addition of between 0.01 and 1.0 μg of viral RNA. After incubation at 34°C for 2 h, aliquots (2 μl) were dissociated in electrophoresis sample buffer and analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis as previously described (8, 23). A mixture of ^{14}C-labeled proteins, supplied by the Radiochemical Centre, was used for the molecular weight standards. For analytical purposes, a gel thickness of 0.7 mm was routinely used, and the labeled proteins were detected by fluorography (24).

For preparative electrophoresis, the gel thickness was increased to 1.5 mm, and the products were detected by direct autoradiography of the dried gel.

Tryptic peptide mapping. Segments from the preparative gels containing products of interest were rehydrated, and the proteins were recovered as previously described (10). The labeled proteins were then oxidized and digested with trypsin, and the resultant peptides were fractionated in two dimensions on cellulose thin-layer plates by following standard procedures (12). Peptide maps of MuMTV structural proteins, the viral protein precursors from infected cells, and the preparation of endo-β-*N*-acetylglucosaminidase H-treated Pr73^{env} have been described in detail elsewhere (8–12).

RESULTS

Translation of different size fractions of MuMTV RNA. Retrovirus genome RNA appears to be highly susceptible to nuclease attack, both within mature virions and during extraction. Thus, even when steps were taken to minimize degradation, such as harvesting virus at 4-h intervals, MuMTV genome RNA was found to be considerably nicked. When the 70S RNA complex was denatured by heating, the resulting single-stranded genome RNA ranged in size from full-length 35S subunits to fragments of about 8 to 10S. Selection of polyadenylated RNAs from this mixture therefore generated a spectrum of molecules of increasing length from the 3′ end. These were fractionated by sucrose gradient centrifugation, and samples from each fraction were tested for their ability to direct cell-free protein synthesis in a rabbit reticulocyte lysate system (Fig. 1). Discrete proteins were produced by all size classes of the viral RNA suggesting that initiation and termination of translation were occurring at specific sites. Moreover, different preparations of viral RNA from both GR and Mm5MT/cl cells gave essentially identical products. Some of these corresponded in size to the precursor polypeptides for the MuMTV structural proteins found in infected cells. For example, RNA of full-genome length directed synthesis of 77,000-, 110,000-, and 160,000-dalton proteins equivalent to the *gag* and *gag-pol* gene products (Fig. 1). This was confirmed by tryptic peptide mapping as discussed below. Thus, initiation and termination were probably occurring at the sites used in vivo for expression of this region of the genome. In contrast, the vast majority of the smaller products were probably a result of either premature termination of translation or initiation at normally cryptic sites made accessible only as a result of degradation of the RNA, or both. Two groups of proteins translated from subgenomic fragments of MuMTV RNA were of particular interest. A series of proteins of molecular weights from about 50,000 to 70,000 daltons were synthesized preferentially by using RNA of about half-genome length (18 to 28S), whereas a second group of proteins in the 18,000- to 36,000-dalton range were the predominant products with RNA smaller than about 18S (Fig. 1). These were examined in more detail as discussed below.

Tryptic peptide mapping of in vitro translation products. To obtain sufficient material for peptide mapping, it was necessary to pool the polyadenylated RNA from gradients of the type shown in Fig. 1 in three major size fractions: (A) 35 to 28S, (B) 28 to 18S, and (C) 18 to 10S. Each size class was then used to program in vitro protein synthesis in rabbit reticulocyte lysates, and the products were fractionated on preparative polyacrylamide gels (Fig. 2). The major [^{35}S]methionine-labeled products, detected by autoradiography as shown in Fig. 2, were eluted and processed for peptide mapping as described elsewhere (10, 12).

(A) 35 to 28S RNA. Previous studies have indicated that translation of genome-length MuMTV RNA yields the Pr77 and Pr110 precursors for the *gag* gene products (8, 28, 42). We have confirmed and extended these observations to include translation of the putative polymerase precursor Pr160 from genome-length RNA. The tryptic peptide maps of each of these polyproteins, synthesized in vitro and labeled with [^{35}S]methionine, are shown in Fig. 3. The fingerprints obtained were similar to those of the corresponding proteins labeled in vivo and indicated that p77, p110, and p160 represent overlapping products from the same region of the MuMTV genome (8, 9). The most likely explanation for this series of products is that they are initiated at a common site and represent continuation of translation beyond the normal termination site at the 3′ end of p77. Thus, in vivo, the largest product Pr160 is not envisaged as being a precursor to Pr110^{gag} and Pr77^{gag}. This was confirmed by kinetic analysis of the in vitro translation of MuMTV genome RNA. As shown in Fig. 4, synthesis of mature p77 was completed several minutes before any p110 was detectable, which in turn preceded p160. Even after prolonged exposure of the autoradiograph to enhance the intensity of the p160 band, complete p160 was not detectable before approximately 30 to 35 min, compared with 18 min for p110 (Fig. 4).

In addition to the expected mature *gag* and *pol* gene products, a number of smaller proteins were also synthesized from RNA in the 35 to 28S fraction. The most prominent of these, p46, clearly contained tryptic peptides characteristic of the *gag* and *pol* gene products (Fig. 3) and

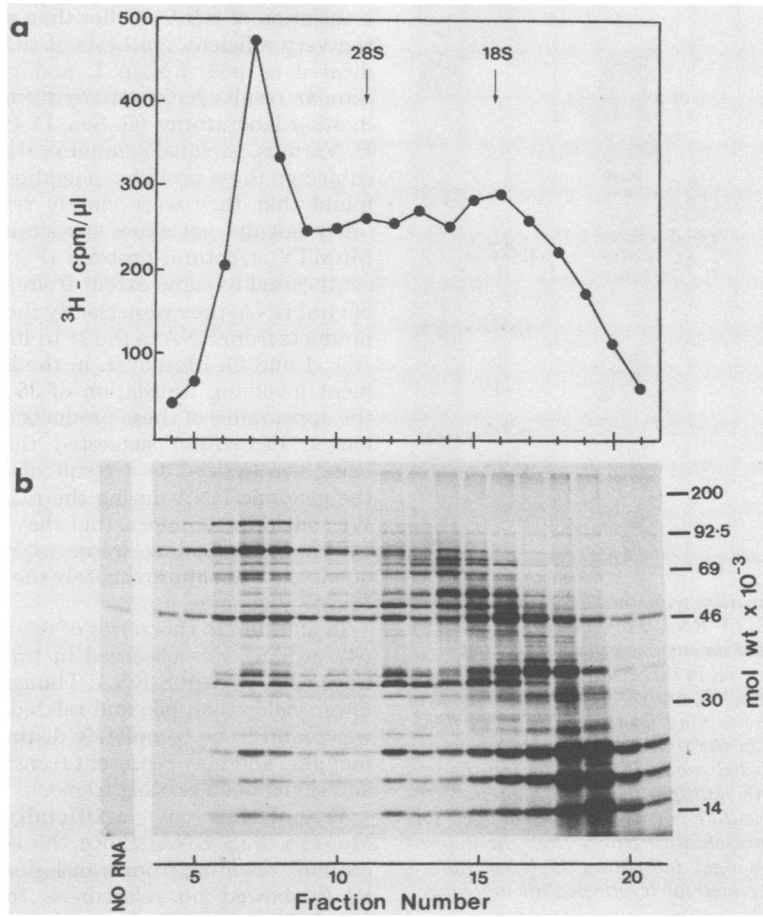

FIG. 1. *Translation of different size fractions of MuMTV polyadenylated RNA. MuMTV 70S RNA, labeled with [^{3}H]uridine, was denatured at 95°C for 2 min, and the polyadenylated RNA molecules were prepared by chromatography on polyuridylic acid-Sepharose (13). (a) Approximately 25 µg of this RNA was fractionated by sucrose gradient centrifugation with ^{32}P-labeled 28 and 18S rRNA's as internal standards as described in the text. Samples of each fraction were counted to determine the distribution of ^{3}H- and ^{32}P-labeled RNA species in the gradient, and the remainder was precipitated twice with ethanol after adding 10 µg of tRNA as a carrier. (b) One-fifth of each fraction was then used to direct in vitro protein synthesis in an RNA-dependent rabbit reticulocyte lysate containing [^{35}S]methionine as label (8, 32). After incubation at 34°C for 2 h, one tenth of each reaction (2 µl) was analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (8, 23). The figure shows an autoradiograph of the ^{33}S-labeled proteins synthesized in response to RNA fractions from the sucrose gradient and analyzed on a 10% acrylamide gel. The track labeled "no RNA" represents the background of protein synthesis in a reticulocyte lysate to which no exogenous RNA was added. The molecular weights of the major products were computed from a series of ^{14}C-labeled protein standards analyzed in a parallel track of the gel as indicated.*

probably results from internal initiation or premature termination, or both. The p34 product (Fig. 2) contained too few methionine-labeled peptides to determine its origin. In some experiments, several other minor products were obtained but these were apparently more efficiently translated from other size fractions of genome RNA. They probably resulted from partial degradation of the RNA in the course of the translation reaction.

(B) 28 to 18S RNA. RNA from the 28 to 18S size fraction, corresponding on average to about half-genome length, directed synthesis of a large number of products (Fig. 1 and 2). Tryptic peptide mapping indicated that many of the larger products were, in fact, related to the *gag* and *pol* precursors produced with fraction A RNA (data not shown), whereas some of the smaller proteins were identical to those synthesized from fraction C RNA. Of the proteins preferentially

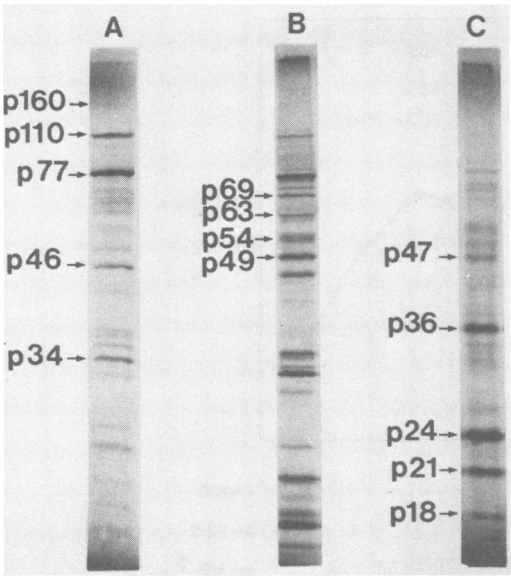

FIG. 2. *Preparative translation of different size fractions of MuMTV RNA. Polyadenylated MuMTV RNA, fractionated on sucrose gradients as shown in Fig. 1, was recovered in three major size classes: (A) 35 to 28S, (B) 28 to 18S, and (C) 18 to 10S. Portions of each fraction, containing up to 1 μg of viral RNA, were translated in the rabbit reticulocyte lysate system. The ^{35}S-labeled products were then purified by electrophoresis in multiple tracks of a preparative sodium dodecyl sulfate-polyacrylamide slab gel. The figure shows representative tracks from the autoradiograph of such a gel, indicating the labeled products which were eluted for tryptic peptide mapping.*

synthesized by using fraction B RNA, the p49, p54, p63, and p69 products shown in Fig. 2 were of particular interest. From peptide mapping, it was clear that these proteins were related to one another and also apparently to the precursor of the MuMTV envelope glycoproteins seen in infected cells (Fig. 5). Since the products of in vitro translation would not be expected to be glycosylated, the peptide maps were compared with that of authentic Pr73^{env} which had been treated with endo-β-*N*-acetylglucosaminidase H to remove the majority of the carbohydrate residues (11) (p61^{env} in Fig. 5). The numbered peptides indicate those which were believed to be specific for the envelope glycoproteins and which were consistently found in these in vitro-translated products. The complexity of the fingerprint of p49 presumably resulted from contamination by a "background" product synthesized from an RNA endogenous to the reticulocyte lysate. This background product was also a contaminant of the p46 in Fig. 2 and 3.

(C) 18 to 10S RNA. A striking feature of the translation of RNA smaller than about 18S was the very efficient synthesis of the products indicated as p36, p24, p21, and p18 in Fig. 2. Similar results have apparently been obtained in other laboratories (G. Sen; D. Robertson and H. Varmus, personal communication). When we subjected these proteins to peptide mapping, we found that they were closely related to each other but did not show any similarities to the MuMTV structural proteins (Fig. 6). Although synthesized to some extent from all size classes of viral RNA, they were clearly the predominant products from RNA in the 18 to 10S size fraction (Fig. 1 and 2). Moreover, in the kinetic experiment involving translation of 35 to 28S RNA, the appearance of these products only after prolonged incubation suggested that they were being synthesized as a result of breakdown of the genomic RNA during the reaction (Fig. 4). We conclude, therefore, that they were encoded by an open reading frame of nucleotide sequences within approximately the 3′ third of the MuMTV genome.

In addition to this series of related proteins, a p47 product was observed in translation reactions with 18 to 10S RNA. Though produced in lower yields than p36 and related proteins, p47 was found to be completely distinct by peptide mapping and may represent translation from an additional open reading frame.

Translation of artificially degraded MuMTV 35S RNA. Since the series of small proteins resulting from translation of 18 to 10S RNA showed no relatedness to any of the MuMTV structural proteins, it was necessary to show that they were definitely encoded by the MuMTV genome RNA and not by a particularly abundant, nonviral mRNA contaminating the viral preparations. Intact, polyadenylated 35S RNA, isolated from sucrose density gradients, was artificially degraded by mild alkali treatment, and fragments of RNA from the 3′ region of the genome were recovered by chromatography on polyuridylic acid-Sepharose. Equivalent amounts of the undegraded and degraded RNA preparations were then used to program in vitro translation reactions. As expected, the predominant products from undegraded 35S RNA were the p77, p110, and p160 *gag* and *pol* precursors and the related p46 *gag* fragment (Fig. 7). After degradation of this RNA, synthesis of the mature *gag* gene products was diminished, and there was a reciprocal enrichment for p24, p21, and p18, indicating that they are indeed encoded by MuMTV genome RNA (Fig. 7).

DISCUSSION

In this study, we have used in vitro translation of MuMTV genome RNA to deduce the order

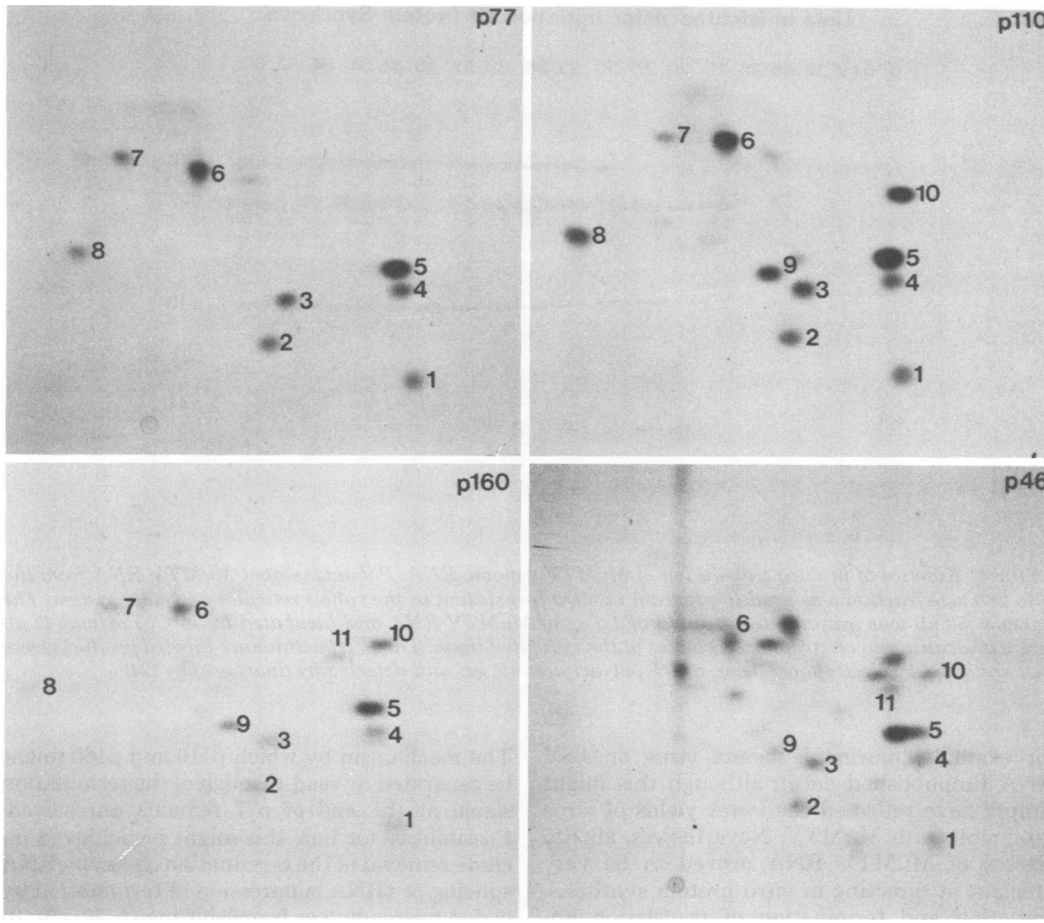

FIG. 3. *Tryptic peptide maps of protein translated from 35 to 28S MuMTV RNA. The p77, p110, p160, and p46 products indicated in Fig. 2A were eluted from the rehydrated gel, oxidized, and digested with trypsin following standard procedures (10, 12). The resultant peptides were fractionated in two dimensions on cellulose thin-layer plates as previously described (12). The first dimension was electrophoresed at pH 4.5 from left to right toward the cathode. The second dimension, from bottom to top, was ascending chromatography. The [^{35}S]methionine-containing peptides were detected by autoradiography with Kodak SB5 X-ray film. The numbered spots on the autoradiographs indicate those peptides which were consistently found in the in vitro translation products and which appeared to correspond to those of the authentic gag gene products (8, 9).*

of the genes coding for the viral structural proteins and to determine whether the genome has the capacity to encode additional, nonstructural proteins whose function may be required for the oncogenic potential of the virus. The rationale behind this approach was that partially degraded viral RNA used to program an in vitro cell-free translation system might allow protein synthesis to start at normally internal methionine codons. If initiation preferentially occurs at the methionine codon closest to the 5′ end of an RNA, it was thought that translation of polyadenylated fragments from randomly degraded viral RNA would permit initiation at all possible methionine codons, in any reading frame, revealing the entire protein-coding capacity of the viral genome. However, other constraints, such as nucleotide sequences in the vicinity of the potential start codon, might affect the initiation of protein synthesis. In addition, primary sequence and secondary structural features might influence the probability of hydrolysis at certain sites on the RNA, leading to distortion of the expected random distribution of RNA fragments. This point may be particularly relevant in the present study in which the natural instability of MuMTV RNA was exploited to generate the subgenomic fragments. In this regard, MuMTV RNA was found to be much more susceptible to breakdown during isolation than,

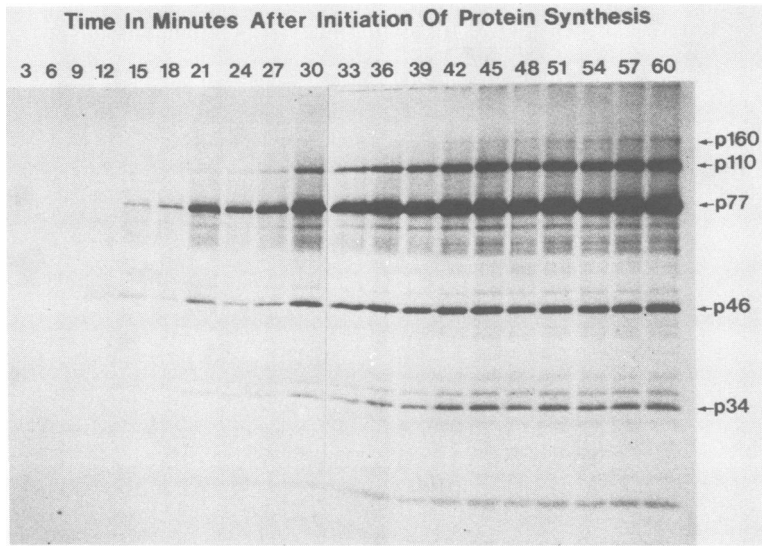

FIG. 4. *Kinetics of in vitro translation of MuMTV genome RNA. Polyadenylated MuMTV RNA from the 35 to 28S size fraction was used to program in vitro translation in the rabbit reticulocyte lysate system. The reaction (50 μl) was initiated by addition of 1.5 μg of MuMTV RNA and incubated at 34°C. Portions (2 μl) were withdrawn into electrophoresis buffer at the indicated times. The [^{35}S]methionine-labeled products were then analyzed by electrophoresis on a 10% polyacrylamide gel and detected by fluorography (24).*

for example, murine leukemia virus or ASV RNA (unpublished data), although this might simply have reflected the lower yields of virus attainable with MuMTV. Nevertheless, all size classes of MuMTV RNA proved to be very efficient at directing in vitro protein synthesis. Initiation and termination of translation appeared to be both specific and reproducible since discrete products were consistently observed which showed preferential synthesis from particular size fractions of the viral RNA (Fig. 1).

Complete genome-length RNA directed the synthesis of p77 and p110, equivalent to the precursors for the MuMTV internal structural proteins, indicating that the *gag* gene is probably located close to the 5′ end of the genome (8, 42). The putative *gag-pol* precursor p160 was also synthesized from this RNA. However, the kinetics of synthesis of these products (Fig. 4) indicated that p160 was not a precursor to the smaller species but that all three were independently translated from the same region of the genome RNA (21). Although we cannot exclude the possibility that these products were translated from different initiation sites on three distinct species of RNA, comparison of their peptide maps with those of the MuMTV structural proteins and intracellular intermediates suggested that they probably had a common amino terminus but different carboxyl termini (10). The mechanism by which p110 and p160 might be generated by read-through of the termination signal at the end of p77 remains unresolved. Possibilities for how this might be achieved include removal of the termination codon by RNA splicing or tRNA suppression of termination by codon misreading or frameshifting (1, 33, 49). In MuMTV, two such events must be envisaged as generating the three overlapping products. Thus, MuMTV appears to conform to the pattern observed in type C retroviruses in that the *gag* and *pol* genes are expressed as separate but overlapping polypeptide precursors (22, 29, 33). This would place the *pol* gene immediately distal to *gag*.

The similarity of the peptide maps of the in vitro products to those of the authentic precursors isolated from infected cells implied that in vitro translation of genome-sized RNA was initiating at the correct site. This was probably not the case for many of the products translated from subgenomic RNA fractions. For example, the series of products p69, p63, p54, and p49, synthesized preferentially from the 28 to 18S RNA fraction, were clearly related in that the tryptic peptides of each species formed a subset of those of the next largest product (Fig. 5). The smallest of these, p49, contained four peptides also common to mature gp36 from MuMTV virions which has been previously shown to be

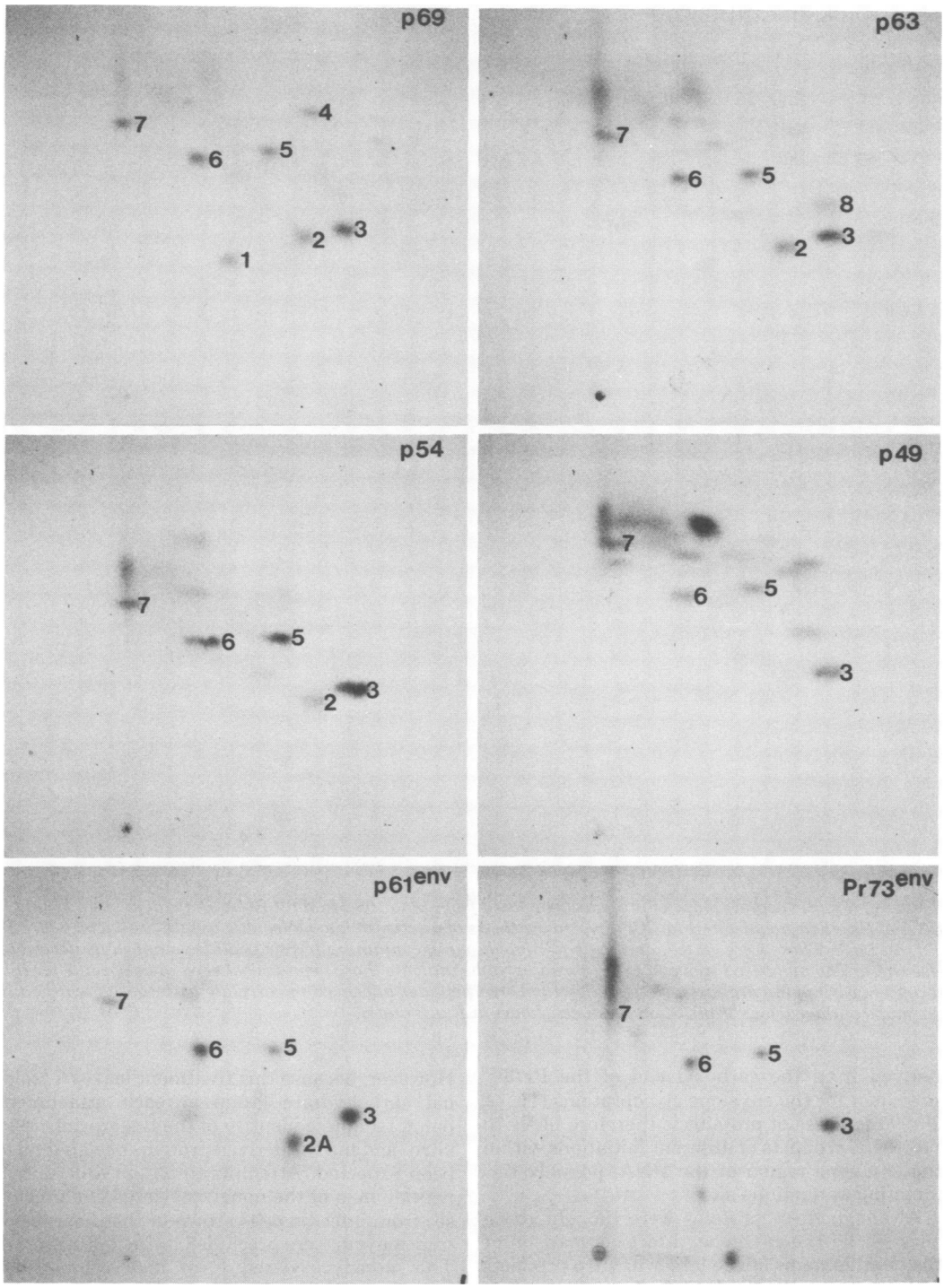

FIG. 5. *Tryptic peptide maps of proteins translated from 28 to 18S MuMTV RNA. The major products obtained from translation of RNA in fraction B (Fig. 2) were eluted and peptide mapped as described in the text. The figure compares the maps of the [^{35}S]methionine-containing peptides from in vitro-synthesized p69, p63, p54, and p49 with those of in vivo-labeled Pr73^{env} and endo-β-N-acetylglucosaminidase H-treated Pr73^{env} (p61^{env}) prepared as previously described (11). The numbered spots on the autoradiographs indicate those peptides shared by the various in vitro- and in vivo-derived products. Peptides 3, 5, and 6 were also found in the mature viral glycoprotein gp36 (12).*

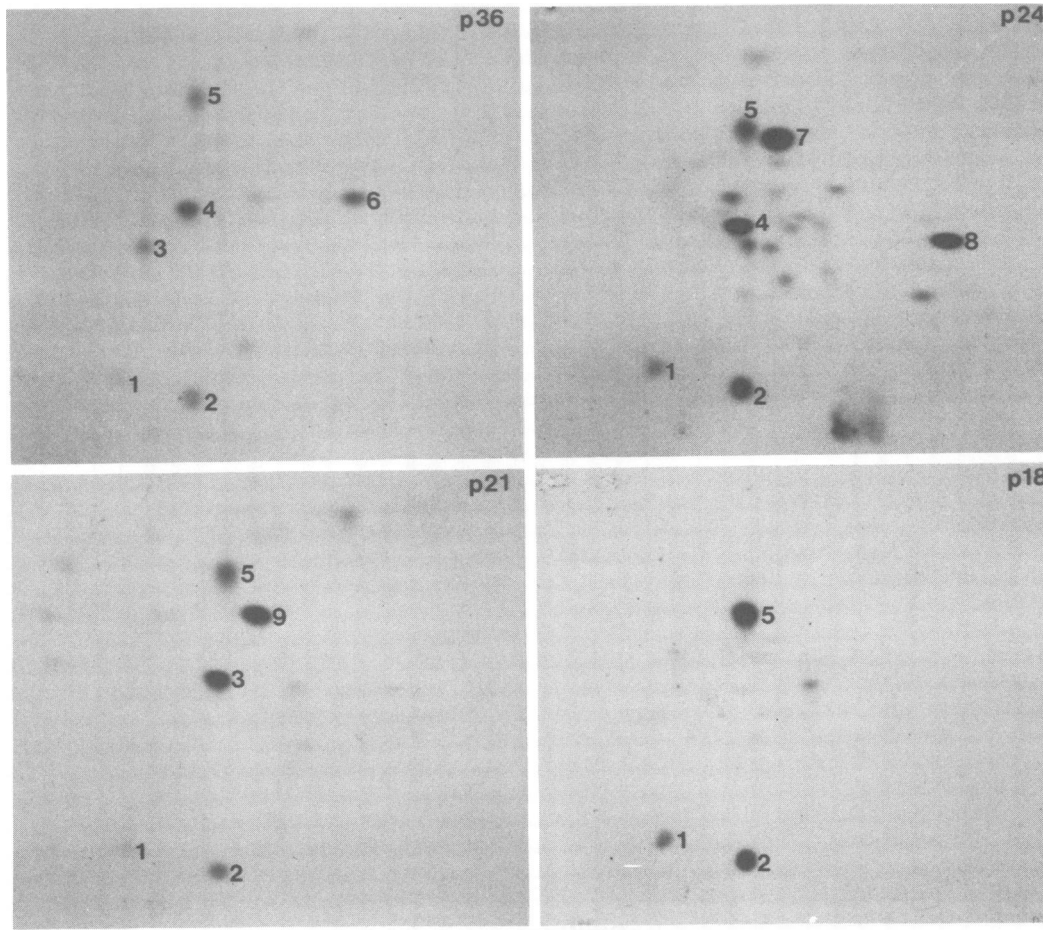

FIG. 6. *Tryptic peptide maps of proteins translated from 18 to 10S MuMTV RNA. Polyadenylated MuMTV RNA of less than about 18S (Fig. 2C) directed synthesis of a series of low-molecular-weight products: p36, p24, p21, and p18. These were eluted, and their [^{35}S]methionine-containing tryptic peptides mapped as described elsewhere. The numbered spots on the autoradiographs indicate those peptides believed to be specific for this series of overlapping products. None of the numbered peptides appeared to correlate with those identified in the maps of the various translation products shown in Fig. 3 and 5.*

derived from the carboxyl end of the Pr73^{env} precursor for the envelope glycoproteins (11, 12, 39). This series of proteins is therefore likely to represent products of different initiations within the *env* gene region of the RNA, possibly terminating at a unique site.

Although these proteins were thought to be encoded by the *env* gene, direct comparison of the peptide maps of the larger in vitro products with that of authentic Pr73^{env} showed only partial similarity, probably as a result of carbohydrate side chains affecting the mobility of some Pr73^{env} peptides. When the majority of the sugars were removed by treatment with endo-β-*N*-acetylglucosaminidase H, the similarity in the peptide maps was more striking (Fig. 5). However, because this treatment leaves a residual carbohydrate group at each attachment point, complete identity of the fingerprints of in vitro- and in vivo-derived proteins need not have been expected. Attempts to obtain a definitive peptide map of the nonglycosylated *env* precursor from infected cells grown in the presence of tunicamycin were precluded by lack of material. This nonglycosylated form had a molecular weight of 60,000, which is considerably smaller than the largest in vitro product, p69 (11). Assuming that no glycosylation had occurred in vitro, then two possible explanations can be envisaged to account for this difference in molecular weights. The first is that p69 was initiated at a site upstream from the normal start of

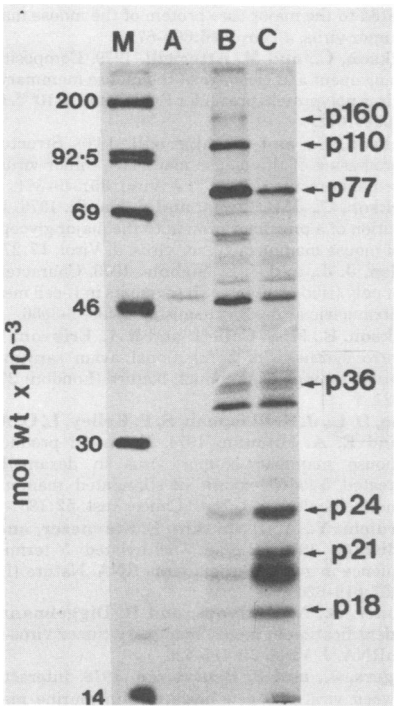

FIG. 7. *Translation of partially degraded MuMTV genome RNA. Full-length, polyadenylated MuMTV RNA was recovered from the 35S region of a gradient of the type shown in Fig. 1. A portion of this RNA was then subjected to milk alkaline hydrolysis as described in the text, and the resultant polyadenylated fragments recovered by polyuridylic acid-Sepharose chromatography. The figure shows an autoradiograph of the [^{35}S]methionine-labeled products synthesized in rabbit reticulocyte lysates with: (A) no exogenous RNA; (B) undegraded MuMTV 35S RNA; (C) polyadenylated fragments from partially degraded MuMTV 35S RNA; (M) the series of ^{14}C-labeled proteins (Radiochemical Centre) used as molecular weight standards. The sizes of the major MuMTV-specific products, designated p160, etc., were computed from their electrophoretic mobilities relative to these standards.*

the *env* gene product. Splicing of the viral RNA in vivo to produce the mRNA for *env* could remove this site and expose the authentic start site. However, in vitro translation of an MuMTV-specific 24S mRNA from infected cells has yielded a 68,000-dalton product (42; D. Robertson, J. Dudley, and H. Varmus, personal communication). This result favors the alternative explanation, that p69 is, in fact, the true primary translation product of the *env* gene, and that processing to remove the signal sequence responsible for transmembrane synthesis of the glycoprotein results in the loss of a 9,000-dalton peptide. Although this leader peptide would be very large compared with other glycoprotein signal sequences (2), a parallel situation appears to exist in the *env* gene product of ASV (E. Hunter, personal communication).

All of the major products of in vitro translation of RNA from 35 to about 18S could generally be attributed to expression of the *gag*, *pol*, and *env* genes. Thus, no potential gene between *pol* and *env* was detected, and synthesis of *env*-related products was optimal from RNA of about half-genome length, suggesting that the gene order was indeed 5′ *gag-pol-env*-3′. However, the possibility still exists that there is an additional gene or genes distal to *env*. Translation of RNA of less than about 18S led to the very efficient synthesis of a series of products whose peptide maps were entirely unrelated to those of the *gag*, *pol*, and *env* gene products (compare Fig. 3, 5, and 6). This group of proteins, the largest of which had a molecular weight of 36,000, again appeared to be a series of overlapping products from the same region of the genome RNA. Since optimal synthesis of the p36 product was obtained with RNA marginally larger than that yielding maximum levels of p24, p21, and p18 (Fig. 1) we concluded that they were probably initiated at different sites on the genome but may have shared a common carboxyl terminus. Preliminary amino acid sequence analysis of the amino terminus of each protein has indicated that all four were initiated with a methionine residue but that they have distinct amino terminal sequences.

Although these small proteins were found to be unrelated to any of the MuMTV structural proteins, they were encoded by the viral genome RNA and not by a contaminating cellular mRNA. This question was resolved by isolating intact, polyadenylated 35S RNA from gradient-purified virus and demonstrating that the ability of this RNA to direct synthesis of this series of products was dependent on partial degradation, presumably to expose normally cryptic initiation sites for translation. Moreover, different preparations of viral RNA from two different cell types gave products with the same molecular weights and peptide maps. If p36 is taken to be the prototype product of a hitherto unidentified gene, the results of in vitro translation imply that expression of this gene would require an mRNA equivalent in size to about 18S rRNA, or about 2,500 nucleotides. The predominant MuMTV-specific mRNA species detected in infected cells have sedimentation coefficients of 35S and 24S, but these analyses do not preclude the existence of minor amounts of a smaller mRNA (17, 37, 42). A 2,500-nucleotide-long mRNA would be sufficient to encode at least 36,000 daltons of protein and at the same time

encompass the 1,200 nucleotides corresponding to the large terminal redundancy at the ends of the MuMTV provirus (43).

We conclude from these studies that the MuMTV genome must contain, near its 3′ end, sequences which provide an open reading frame for translation of at least 36,000 daltons of protein. DNA sequencing data from other eucaryotic viruses and cloned cellular genes have indicated that it is unusual for there to be a substantial stretch of nucleotide sequence with an open reading frame which is not expressed in a functional protein. It is therefore conceivable that p36 represents all or part of a potentially functional gene product encoded by MuMTV. Analogous studies on the translation of ASV have permitted in vitro synthesis of a functional, transformation-specific protein indistinguishable from its counterpart in transformed cells (14, 41). Although it is unlikely that an equivalent protein is specified by the MuMTV genome, it is still feasible that a virally coded, nonstructural protein may play a role in tumorigenesis by increasing the probability that an infected mammary epithelial cell will respond to some other oncogenic stimulus. A test for a functional role for the p36 protein will be whether or not it or a related protein can be detected in infected cells or in mouse mammary tumors.

ACKNOWLEDGMENTS

We thank Maureen Atterwill and Lynne Gilbert for excellent technical assistance and Harold Varmus, Paula Enrietto, and Michael Hayman for critical reading of the manuscript. We also thank Audrey Gibson for help in preparing the manuscript.

LITERATURE CITED

1. **Atkins, J. F., R. F. Gesteland, B. R. Reid, and C. W. Anderson.** 1979. Normal tRNAs promote ribosomal frameshifting. Cell **18**:1119–1131.
2. **Austen, B. M.** 1979. Predicted secondary structures of amino-terminal extension sequences of secreted proteins. FEBS Lett. **103**:308–313.
3. **Baltimore, D.** 1974. Tumor viruses: 1974. Cold Spring Harbor Symp. Quant. Biol. **39**:1187–1200.
4. **Brugge, J. S., and R. L. Erikson.** 1977. Identification of a transformation specific antigen induced by an avian sarcoma virus. Nature (London) **269**:346–348.
5. **Coffin, J. M., and M. A. Billeter.** 1976. A physical map of the Rous sarcoma virus genome. J. Mol. Biol. **100**: 293–318.
6. **Collett, M. S., and R. L. Erikson.** 1978. Protein kinase activity associated with the avian sarcoma virus *src* gene product. Proc. Natl. Acad. Sci. U.S.A. **75**:2021–2024.
7. **Collett, M. S., E. Erikson, A. F. Purchio, J. S. Brugge, and R. L. Erikson.** 1979. A normal cell protein similar in structure and function to the avian sarcoma virus transforming gene product. Proc. Natl. Acad. Sci. U.S.A. **76**:3159–3163.
8. **Dahl, H.-H. M., and C. Dickson.** 1979. Cell-free synthesis of the mouse mammary tumor virus Pr77 from virion and intracellular mRNA. J. Virol. **29**:1131–1141.
9. **Dickson, C., and M. Atterwill.** 1978. Polyproteins related to the major core protein of the mouse mammary tumor virus. J. Virol. **26**:660–672.
10. **Dickson, C., and M. Atterwill.** 1979. Composition, arrangement and cleavage of the mouse mammary tumor virus polyprotein precursor $Pr77^{gag}$ and $p110^{gag}$. Cell **17**: 1003–1012.
11. **Dickson, C., and M. Atterwill.** 1980. Structure and processing of the mouse mammary tumor virus glycoprotein precursor $Pr73^{env}$. J. Virol. **35**:349–361.
12. **Dickson, C., J. P. Puma, and S. Nandi.** 1976. Identification of a precursor protein to the major glycoproteins of mouse mammary tumor virus. J. Virol. **17**:275–282.
13. **Eiden, J. J., and J. L. Nichols.** 1973. Characterization of poly (riboadenylic acid) segments in L-cell messenger ribonucleic acids. Biochemistry **12**:3951–3956.
14. **Erikson, E., M. S. Collett, and R. L. Erikson.** 1978. In vitro synthesis of a functional avian sarcoma virus transforming-gene product. Nature (London) **274**:919–921.
15. **Fine, D. L., J. K. Plowman, S. P. Kelley, L. O. Arthur, and E. A. Hillman.** 1974. Enhanced production of mouse mammary tumor virus in dexamethasone-treated 5-iododeoxyuridine-stimulated mammary tumor cell cultures. J. Natl. Cancer Inst. **52**:1881–1886.
16. **Furuichi, Y., A. J. Shatkin, E. Stavnezer, and J. M. Bishop.** 1975. Blocked methylated 5′-terminal sequence in avian sarcoma virus RNA. Nature (London) **257**:618–620.
17. **Groner, B., N. E. Hynes, and H. Diggelmann.** 1979. Identification of mouse mammary tumor virus-specific mRNA. J. Virol. **30**:417–420.
18. **Hilgers, J., and P. Bentvelzen.** 1978. Interaction between viral and genetic factors in murine mammary cancer. Adv. Cancer Res. **26**:143–195.
19. **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.
20. **Keith, J., and H. Fraenkel-Conrat.** 1975. Identification of the 5′ end of Rous sarcoma virus RNA. Proc. Natl. Acad. Sci. U.S.A. **72**:3347–3350.
21. **Kerr, I. M., U. Olshevsky, H. F. Lodish, and D. Baltimore.** 1976. Translation of murine leukemia virus RNA in cell-free systems from animal cells. J. Virol. **18**: 627–635.
22. **Kopchick, J. J., G. A. Jamjoom, K. F. Watson, and R. B. Arlinghaus.** 1978. Biosynthesis of reverse transcriptase from Rauscher murine leukemia virus by synthesis and cleavage of a gag-pol read-through viral precursor polyprotein. Proc. Natl. Acad. Sci. U. S. A. **75**:2016–2020.
23. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
24. **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.
25. **Levinson, A. D., H. Oppermann, L. Levintow, H. E. Varmus, and J. M. Bishop.** 1978. Evidence that the transforming gene of avian sarcoma virus encodes a protein kinase associated with a phosphoprotein. Cell **15**:561–572.
26. **Martin, G. S.** 1970. Rous sarcoma virus: a function required for maintenance of the transformed state. Nature (London) **227**:1021–1029.
27. **Nandi, S., and C. M. McGrath.** 1973. Mammary neoplasia in mice. Adv. Cancer Res. **17**:353–414.
28. **Nusse, R., F. A. M. Asselbergs. M. H. L. Salden, R. J. A. M. Michalides, and H. Bloemendal.** 1978. Translation of mouse mammary tumor virus RNA: precursor polypeptides are phosphorylated during processing. Virology **91**:106–115.
29. **Oppermann, H., J. M. Bishop, H. E. Varmus, and L.**

Levintow. 1977. A joint product of the genes *gag* and *pol* of avian sarcoma virus: a possible precursor of reverse transcriptase. Cell **12**:993-1005.

30. **Oppermann, H., A. D. Levinson, H. E. Varmus, L. Levintow, and J. M. Bishop.** 1979. Uninfected vertebrate cells contain a protein that is closely related to the product of the avian sarcoma virus transforming gene (*src*). Proc. Natl. Acad. Sci. U.S.A. **76**:1804-1808.
31. **Papkoff, J., T. Hunter, and K. Beemon.** 1980. *In vitro* translation of virion RNA from Moloney murine sarcoma virus. Virology **101**:91-103.
32. **Pelham, H. R. B., and R. J. Jackson.** 1976. An efficient mRNA-dependent translation system from reticulocyte lysates. Eur. J. Biochem. **67**:247-256.
33. **Philipson, L., P. Anderson, U. Olshevsky, R. Weinberg, D. Baltimore, and R. Gesteland.** 1978. Translation of murine leukemia and sarcoma virus RNAs in nuclease-treated reticulocyte extracts: enhancement of the *gag-pol* polypeptide with yeast suppressor tRNA. Cell **13**:189-199.
34. **Racevskis, J., and N. H. Sarkar.** 1978. Synthesis and processing of precursor polypeptides to murine mammary tumor virus structural proteins. J. Virol. **25**:374-383.
35. **Ringold, G., E. Y. Lasfargues, J. M. Bishop, and H. E. Varmus.** 1975. Production of mouse mammary tumor virus by cultured cells in the absence and presence of hormones: assay by molecular hybridization. Virology **65**:135-147.
36. **Ringold, G. M., K. R. Yamamoto, P. R. Shank, and H. E. Varmus.** 1977. Mouse mammary tumor virus DNA in infected rat cells: characterization of unintegrated forms. Cell **10**:19-26.
37. **Robertson, D. L., and H. E. Varmus.** 1979. Structural analysis of the intracellular RNAs of murine mammary tumor virus. J. Virol. **30**:576-589.
38. **Schochetman, G., C. W. Long, S. Oroszlan, L. Arthur, and D. L. Fine.** 1978. Isolation of separate precursor polypeptides for mouse mammary tumor virus glycoproteins and nonglycoproteins. Virology **85**:168-174.
39. **Schochetman, G., S. Oroszlan, L. Arthur, and D. L. Fine.** 1977. Gene order of the mouse mammary tumor virus glycoproteins. Virology **83**:72-83.
40. **Sefton, B. M., K. Beemon, and T. Hunter.** 1978. Comparison of the expression of the *src* gene of Rous sarcoma virus in vitro and in vivo. J. Virol. **28**:957-971.
41. **Sefton, B. M., T. Hunter, and K. Beemon.** 1979. Product of in vitro translation of Rous sarcoma virus *src* gene has protein kinase activity. J. Virol. **30**:311-318.
42. **Sen, G. C., S. W. Smith, S. L. Marcus, and N. H. Sarkar.** 1979. Identification of the messenger RNAs coding for the *gag* and *env* gene products of the murine mammary tumor virus. Proc. Nat. Acad. Sci. U.S.A. **76**: 1736-1740.
43. **Shank, P. R., J. C. Cohen, H. E. Varmus, K. R. Yamamoto, and G. M. Ringold.** 1978. Mapping of linear and circular forms of mouse mammary tumor virus DNA with restriction endonucleases: evidence for a large specific deletion occurring at high frequency during circularization. Proc. Natl. Acad. Sci. U.S.A. **75**: 2112-2116.
44. **Shields, A., S. Goff, M. Paskind, G. Otto, and D. Baltimore.** 1979. Structure of the Abelson murine leukemia virus genome. Cell **18**:955-962.
45. **Tooze, J. (ed.).** 1973. The molecular biology of tumour viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
46. **Varmus, H. E., G. Ringold, and K. R. Yamamoto.** 1979. Regulation of mouse mammary tumor virus gene expression by glucocorticoid hormones, p. 254-278. *In* J. D. Baxter and G. G. Rousseau (ed.), Glucocorticoid hormone action. Springer-Verlag, New York.
47. **Wang, L.-H., and P. Duesberg.** 1974. Properties and location of poly(A) in Rous sarcoma virus RNA. J. Virol. **14**:1515-1529.
48. **Wang, L.-H., D. Galehouse, P. Mellon, P. Duesberg, W. S. Mason, and P. K. Vogt.** 1976. Mapping oligonucleotides of Rous sarcoma virus RNA that segregate with polymerase and group-specific antigen markers in recombinants. Proc. Natl. Acad. Sci. U.S.A. **73**:3952-3956.
49. **Weiss, S. R., P. B. Hackett, H. Oppermann, A. Ullrich, L. Levintow, and J. M. Bishop.** 1978. Cell-free translation of avian sarcoma virus RNA. Suppression of the *gag* termination codon does not augment synthesis of the joint *gag-pol* product. Cell **15**:607-614.
50. **Zavada, J., C. Dickson, and R. Weiss.** 1977. Pseudotypes of vesicular stomatitis virus with envelope antigens provided by murine mammary tumor virus. Virology **82**:221-231.